

SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB#

21728

RECEIVED

CRFE

Requester's Full Name: Samuel Wei Liu Examiner #: 79120 Date: 7-24-2002
 Art Unit: 1653 Phone Number 306-3483 Serial Number: 09600787
 Mail Box and Bldg/Room Location: 9B01 Results Format Preferred (circle): PAPER DISK E-MAIL
 (9D08)

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): Jarman, C. D.

Earliest Priority Filing Date: 1-22-1998

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

please search SEQ ID NO: 1 (open & close)

Thanks.

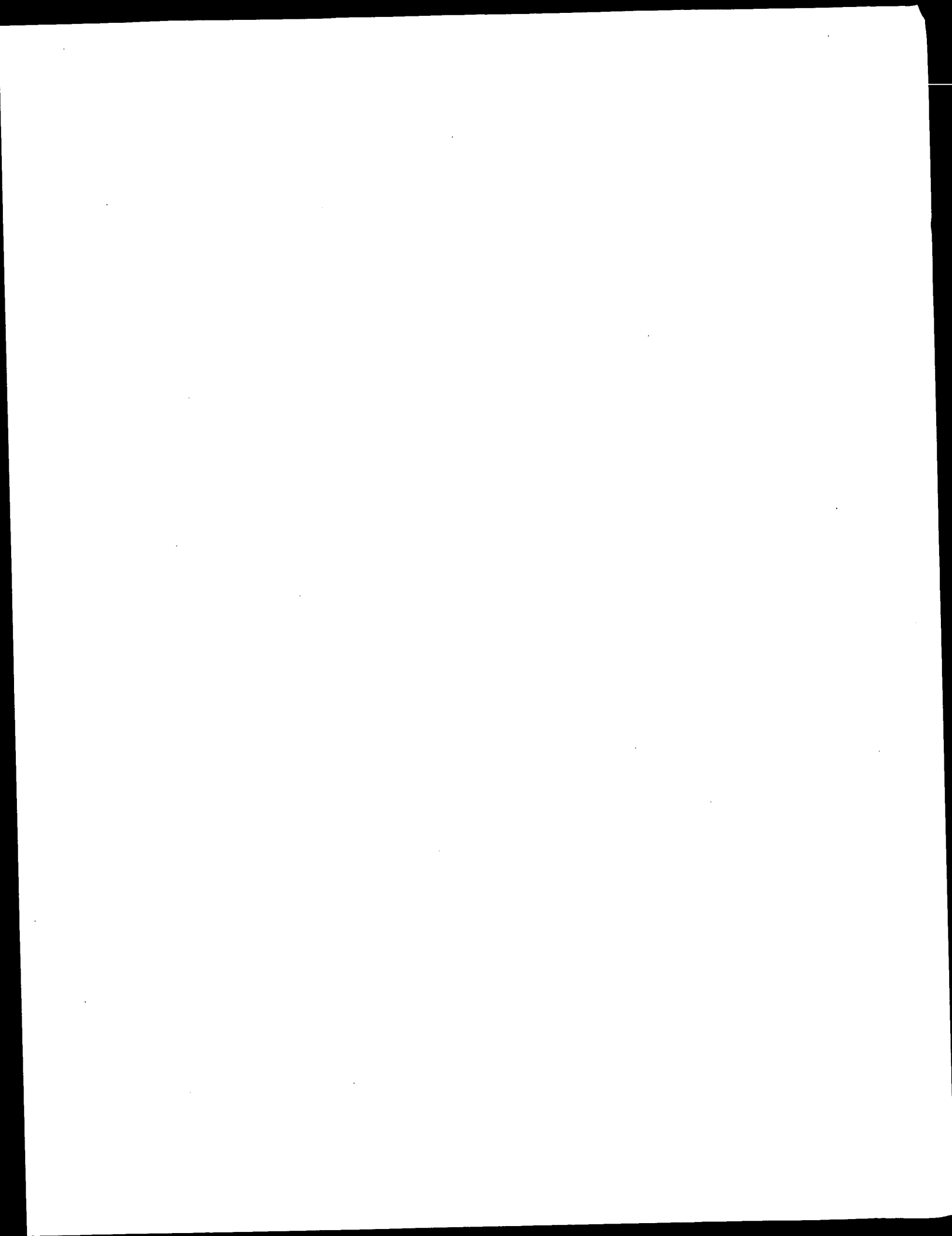
STAFF USE ONLY

Searcher: _____ Point of Contact: Alexandra Wacław
 Searcher Phone #: _____ Technical Info. Specialist
 Searcher Location: CM1 6A02 Tel: 308-4491
 Date Searcher Picked Up: 7-29-02
 Date Completed: 7-29-02
 Searcher Prep & Review Time: 4
 Clerical Prep Time: _____
 Online Time: 6

Type of Search

NA Sequence (#) _____ STN _____
 AA Sequence (#) 1 Dialog _____
 Structure (#) _____ Questel/Orbit _____
 Bibliographic _____ Dr. Link _____
 Litigation _____ Lexis/Nexis _____
 Fulltext _____ Sequence Systems _____
 Patent Family _____ WWW/Internet _____
 Other _____ Other (specify) Compukey

Vendors and cost where applicable



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2002, 06:47:51 ; Search time 24.8 Seconds

(without alignments)
823.121 Million cell updates/sec

Title: US-09-600-787-1

Perfect score: 610
Sequence: 1 DEOPTIGSGNNVTGSGSKN.....NTVSGSNHVSGSNKVVTDA 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

SPTRMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.prodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	610	100.0	118	10 Q9M3W4	Q9M3W4 <i>loium pere</i>
2	192.5	31.6	922	2 Q9L964	Q9L964 <i>moraxella c</i>
3	172	28.2	892	2 Q9KX38	Q9KX38 <i>moraxella c</i>
4	171	28.0	894	2 Q9L962	Q9L962 <i>moraxella c</i>
5	169.5	27.8	941	2 Q9XD54	Q9XD54 <i>moraxella c</i>
6	166	27.2	889	2 Q9L961	Q9L961 <i>moraxella c</i>
7	155.5	25.5	912	2 Q9XD52	Q9XD52 <i>moraxella c</i>
8	153.5	25.2	863	2 Q9XD56	Q9XD56 <i>moraxella c</i>
9	150.5	24.7	873	2 Q9L960	Q9L960 <i>moraxella c</i>
10	143.5	23.5	560	5 Q61085	Q61085 <i>dictyosteli</i>
11	142	23.3	718	5 Q9XTP6	Q9XTP6 <i>dictyosteli</i>
12	141	23.1	1176	12 Q9S87	Q9S87 <i>paramocium</i>
13	140.5	23.0	832	2 Q54356	Q54356 <i>moraxella c</i>
14	139.5	22.9	1650	5 Q77328	Q77328 <i>plasmodium</i>
15	134.5	22.0	1275	5 Q76602	Q76602 <i>caenorhabdi</i>
16	133	21.8	337	5 Q9W4M1	Q9W4M1 <i>drosophila</i>

17	132.5	21.7	373	5 Q9U217	Q9U217 <i>caenorhabdi</i>
18	131.5	21.6	1096	5 Q97257	Q97257 <i>plasmodium</i>
19	131	21.5	343	10 Q81761	Q81761 <i>arabidopsis</i>
20	129.5	21.2	513	5 Q61067	Q61067 <i>plasmodium</i>
21	129.5	21.2	1001	3 Q05164	Q05164 <i>saccharomyc</i>
22	129.5	21.2	1318	5 Q95P44	Q95P44 <i>dictyosteli</i>
23	129	21.1	503	5 Q96563	Q96563 <i>plasmodium</i>
24	128.5	21.1	96	5 Q9Y1D5	Q9Y1D5 <i>plasmodium</i>
25	128.5	21.1	1043	5 Q95P43	Q95P43 <i>dictyosteli</i>
26	128	21.0	608	5 Q94474	Q94474 <i>dictyosteli</i>
27	128	21.0	1844	5 Q97287	Q97287 <i>plasmodium</i>
28	127.5	20.9	954	5 Q9XYL1	Q9XYL1 <i>dictyosteli</i>
29	127	20.8	604	5 Q26021	Q26021 <i>plasmodium</i>
30	127	20.8	1560	5 Q9GKX5	Q9GKX5 <i>dictyosteli</i>
31	127	20.8	1561	5 Q9GKX5	Q9GKX5 <i>dictyosteli</i>
32	127	20.8	1570	5 Q9U0H8	Q9U0H8 <i>plasmodium</i>
33	124.5	20.4	1436	5 Q96158	Q96158 <i>plasmodium</i>
34	123.5	20.2	609	5 Q9NGW8	Q9NGW8 <i>dictyosteli</i>
35	123.5	20.2	1256	5 Q25770	Q25770 <i>plasmodium</i>
36	123.5	20.2	1278	5 Q96228	Q96228 <i>plasmodium</i>
37	123	20.2	817	5 Q964V6	Q964V6 <i>dictyosteli</i>
38	122.5	20.1	2150	5 Q23863	Q23863 <i>dictyosteli</i>
39	122	20.0	1181	5 Q97260	Q97260 <i>plasmodium</i>
40	121.5	19.9	109	5 Q9Y1D4	Q9Y1D4 <i>plasmodium</i>
41	121.5	19.9	1064	5 Q25773	Q25773 <i>plasmodium</i>
42	121.5	19.9	1231	5 Q97140	Q97140 <i>dictyosteli</i>
43	121.5	19.9	1529	5 Q9GQC2	Q9GQC2 <i>dictyosteli</i>
44	121	19.8	565	5 Q15754	Q15754 <i>dictyosteli</i>
45	121	19.8	1208	5 Q97101	Q97101 <i>dictyosteli</i>
46	120.5	19.8	1543	5 Q9GV71	Q9GV71 <i>dictyosteli</i>
47	120	19.7	234	5 Q25769	Q25769 <i>plasmodium</i>
48	119.5	19.6	682	5 Q9GPR8	Q9GPR8 <i>dictyosteli</i>
49	119.5	19.6	3933	5 Q97239	Q97239 <i>plasmodium</i>
50	119	19.5	559	5 Q9U3Y8	Q9U3Y8 <i>plasmodium</i>
51	119	19.5	967	3 Q08294	Q08294 <i>saccharomyc</i>
52	119	19.5	1546	5 Q95P11	Q95P11 <i>dictyosteli</i>
53	119	19.5	1997	5 Q96239	Q96239 <i>plasmodium</i>
54	118.5	19.4	239	5 Q25772	Q25772 <i>plasmodium</i>
55	118.5	19.4	1755	17 Q26812	Q26812 <i>methanother</i>
56	118	19.3	686	5 Q96245	Q96245 <i>plasmodium</i>
57	117.5	19.3	102	5 Q9U420	Q9U420 <i>plasmodium</i>
58	117	19.2	655	5 Q26109	Q26109 <i>plasmodium</i>
59	117	19.2	2035	2 Q9XCJ4	Q9XCJ4 <i>salmonella</i>
60	116.5	19.1	84	5 Q9Y1F6	Q9Y1F6 <i>plasmodium</i>
61	116.5	19.1	955	5 Q97238	Q97238 <i>plasmodium</i>
62	116	19.0	109	5 Q9U419	Q9U419 <i>plasmodium</i>
63	116	19.0	1213	5 Q95P42	Q95P42 <i>dictyosteli</i>
64	116	19.0	1686	5 Q95P45	Q95P45 <i>dictyosteli</i>
65	116	19.0	2240	5 Q97298	Q97298 <i>plasmodium</i>
66	115.5	18.9	232	5 Q9BHT5	Q9BHT5 <i>plasmodium</i>
67	115.5	18.9	699	5 Q15816	Q15816 <i>dictyosteli</i>
68	115.5	18.9	1360	5 Q94649	Q94649 <i>plasmodium</i>
69	115.5	18.9	2112	5 Q9YEL9	Q9YEL9 <i>drosophila</i>
70	115.5	18.9	2500	5 Q96223	Q96223 <i>plasmodium</i>
71	115.5	18.9	2742	5 Q15801	Q15801 <i>plasmodium</i>
72	115	18.9	97	5 Q9Y1D6	Q9Y1D6 <i>plasmodium</i>
73	115	18.9	115	5 Q9Y1D2	Q9Y1D2 <i>plasmodium</i>
74	115	18.9	510	5 Q97251	Q97251 <i>plasmodium</i>
75	115	18.9	757	5 Q96S66	Q96S66 <i>dictyosteli</i>
76	115	18.9	1136	2 Q9XKQ3	Q9XKQ3 <i>salmonella</i>
77	115	18.9	1963	2 Q9XKQ3	Q9XKQ3 <i>salmonella</i>
78	114.5	18.8	92	5 Q9Y1E6	Q9Y1E6 <i>plasmodium</i>
79	114.5	18.8	728	5 Q96422	Q96422 <i>plasmodium</i>
80	114	18.7	115	5 Q9U417	Q9U417 <i>plasmodium</i>
81	114	18.7	515	5 Q15912	Q15912 <i>dictyosteli</i>
82	114	18.7	709	5 Q16783	Q16783 <i>caenorhabdi</i>
83	114	18.7	720	3 Q9P3M2	Q9P3M2 <i>neurospora</i>
84	114	18.7	1004	16 Q9P6D3	Q9P6D3 <i>xyella fas</i>
85	113.5	18.6	325	5 Q94492	Q94492 <i>dictyosteli</i>
86	113.5	18.6	1649	16 Q9CFA2	Q9CFA2 <i>lactococcus</i>
87	113	18.5	534	5 Q43987	Q43987 <i>dictyosteli</i>
88	113	18.5	2123	5 Q9U9S7	Q9U9S7 <i>dictyosteli</i>
89	112.5	18.4	714	10 Q9LS95	Q9LS95 <i>arabidopsis</i>

90 112 18.4 3978 5 097236 097236 plasmodium
91 111 18.2 247 5 09V4F0 09V4F0 drosophila
92 111 18.2 1736 5 095PH7 095PH7 dictyosteli
93 110.5 18.1 90 5 09YIE1 09YIE1 plasmodium
94 110.5 18.1 368 5 09VRE1 09VRE1 drosophila
95 110.5 18.1 568 5 09NLR3 09NLR3 pinctada ma
96 110.5 18.1 645 5 09VX55 09VX55 drosophila
97 110.5 18.1 800 5 043988 043988 dictyosteli
98 110 18.0 91 5 09YIE1 09YIE1 plasmodium
99 110 18.0 91 5 09YIE4 09YIE4 plasmodium
100 110 18.0 669 5 025771 025771 plasmodium

ALIGNMENTS

RESULT 1
ID 09M3M4 PRELIMINARY; PRT; 118 AA.

AC 09M3M4; PRELIMINARY; PRT; 118 AA.
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE ICE RECRYSTALLISATION INHIBITION PROTEIN (FRAGMENT).
OS Lolium perenne (Perennial ryegrass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Poaceae; Lolium.
OX NCBI_TaxID=4522;
RN [1]
RP SEQUENCE FROM N.A.
RA Sidebottom C.M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RT [2]
RP SEQUENCE FROM N.A.
RA Sidebottom C.M.;
RL Thesis (1999), University of York.
DR EMBL; AJ277399; CAB87814.1; -.
FT NON TER 1 1
SQ SEQUENCE 118 AA; 11766 MW; 0362C1E4F98AE9F8 CRC64;

Query Match 100.0%; Score 610; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 2.7e-35;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DEOPTISGNNVTSSGSKNYLAGNDNTVSGDNNVSGSNNVTGSGNDNTVTSNHYVS 60
DB 1 DEOPTISGNNVTSSGSKNYLAGNDNTVSGDNNVSGSNNVTGSGNDNTVTSNHYVS 60
OY 61 GTNHIVTDNNNNVSGDNNVSGSFHTVSGGNTVSGSNNVTGSGNHVYSGSNKYVTD 118
DB 61 GTNHIVTDNNNNVSGDNNVSGSFHTVSGGNTVSGSNNVTGSGNHVYSGSNKYVTD 118

RESULT 2
ID 09L964 PRELIMINARY; PRT; 922 AA.
AC 09L964;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
USPAL.
GN Moraxella catarrhalis.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=012E;
RX MEDLINE=20138164; PubMed=10671460;
LaFontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr.,
Hansen E.J.;

RT "The UsPAL Protein and a Second Type of UsPA2 Protein Mediate
RT Adherence of Moraxella catarrhalis to Human Epithelial Cells in
RT Vitro.";
RT J. Bacteriol. 182:1364-1373(2000).
DR EMBL; AF181072; AAF40118.1; -.
DR InterPro; IPR000515; BPD.transp.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1
SQ SEQUENCE 922 AA; 96966 MW; A4EC3BDA3973535A CRC64;

Query Match 31.6%; Score 192.5; DB 2; Length 922;
Best Local Similarity 35.9%; Pred. No. 7.3e-06;
Matches 42; Conservative 18; Mismatches 48; Indels 9; Gaps 1;

OY 5 NTISGNNVTSSGSKNYLAGNDNTVSGDNNVSGSNNVTGSGNDNTVTSNHYVSGTNIH 64
DB 146 NTAKGTHSTVIGGRKRNABGNSYVAGGNDNATGNNSTVAGGSKNATGAGSFAGVEN 205
OY 65 IYTDNNNNVSGDNNVSGSFHTVSGGNTVSGSNNVTGSGNHVYSGSNKYVTD 112
DB 206 QAKTEAVALGKNTIGTNSVVAIGSNNVTEDGKODVFLIGSNTTMAOSSVLLGNN 262

RESULT 3
ID 09KX38 PRELIMINARY; PRT; 892 AA.

AC 09KX38; PRELIMINARY; PRT; 892 AA.
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE USPAL.
GN USPAL.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=046E;
RX MEDLINE=20138164; PubMed=10671460;
LaFontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr.,
Hansen E.J.;
RT "The UsPAL Protein and a Second Type of UsPA2 Protein Mediate
RT Adherence of Moraxella catarrhalis to Human Epithelial Cells in
RT Vitro.";
RT J. Bacteriol. 182:1364-1373(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=046E;
RX MacIver I., Latimer J.L., Cope L.D., Thomas S.E., Hansen E.J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U61725; AAF36416.1; -.
DR InterPro; IPR000515; BPD.transp.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
SQ SEQUENCE 892 AA; 93358 MW; 9D71A369672F44C7 CRC64;

Query Match 28.2%; Score 172; DB 2; Length 892;
Best Local Similarity 32.7%; Pred. No. 0.00018;
Matches 35; Conservative 19; Mismatches 47; Indels 6; Gaps 1;

OY 9 GSNVTSSGSKNYLAGNDNTVSGDNNVSGSNNVTGSGNDNTVTSNHYVSGTNIHVT 68
DB 143 GDSSTIGGTYNATGKSTVAGGRNNGATGNNSTVAGGSGYNGATGNNSTVAGGSH---- 198
OY 69 NNNNVSGDNNVSGSFHTVSGGNTVSGSNNVTGSGNHVYSGSNKYV 115
DB 199 --NQATGEGSFAGVENKANNNAVALGKNTIDGNSVVAIGSNNNTI 243

RESULT 4
ID 09L962 PRELIMINARY; PRT; 894 AA.
AC 09L962;

DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE USPA2H.
 GN Moraxella catarrhalis.
 OS Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OC NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O46E;
 RX MEDLINE=20138164; PubMed=10671460;
 RA Lafontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr., Hansen E.J.;
 RT "The Uspl1 Protein and a Second Type of Uspl2 Protein Mediate Adherence of Moraxella catarrhalis to Human Epithelial Cells in Vitro."
 RL J. Bacteriol. 182:1364-1373(2000).
 DR EMBL: AF181074; AAF40120.1; -
 SQ SEQUENCE 894 AA; 94268 MW; 84BFE5DA1CCF14BC CRC64;

Query Match 28.0%; Score 171; DB 2; Length 894;
 Best Local Similarity 34.7%; Pred. No. 0.00021;
 Matches 43; Conservative 19; Mismatches 50; Indels 12; Gaps 4;

OY 5 NTIGSNNVTYRSGSKNVLAGNDNTVIGDNNVSGSNNVTYVGSNNHVVSGT-N 63
 DB 104 NOAKGEHSTIAGGKNQATGRNSTVAGSNNQAVGTINSTVAGSNNQAKNSFAGVGN 163
 OY 64 HIVTDN-----NNVSGNNNNVSGSFHTVSGGHNTV--SGSNNTVSGSNHVVSG--SN 112
 DB 164 QANDNNAVALGKNNTINGNSAIGSENTYVENOKNFFILGSSNTTNAQSSVLLGHETSG 223
 OY 113 KVT 116
 DB 224 KEAT 227

RESULT 5
 ID 09XD54 PRELIMINARY; PRT; 941 AA.
 AC 09XD54;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE USPA1.
 GN USPA1.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OC NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TTA24;
 RX MEDLINE=99315796; PubMed=10383971;
 RA Cope L.D., Lafontaine E.R., Slaughter C.A., Hasemann C.A. Jr., Aebi C., Henderson F.W., McCracken G.H. Jr., Hansen E.J.;
 RT "Characterization of the Moraxella catarrhalis uspl1 and uspl2 genes and their encoded products."
 RL J. Bacteriol. 181:4026-4034(1999).
 DR EMBL: AF113608; AAD43467.1; -
 DR InterPro: IPR000515; BPD_transp.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 SQ SEQUENCE 941 AA; 99079 MW; F5ED65636F474AF CRC64;

Query Match 27.8%; Score 169.5; DB 2; Length 941;
 Best Local Similarity 35.8%; Pred. No. 0.00028;
 Matches 43; Conservative 16; Mismatches 50; Indels 9; Gaps 4;

OY 5 NTIGSNNVTYRSGSKNVLAGNDNTVIGDNNVSGSNNVTYVGSNNHVVSGT-N 63
 DB 108 NEAIGKNTSTVGGFJTNEMGEYSTVAGANNQAKNSYTVYGGGNGNAIGNSTVGGSN 167

OY 64 HIVTDNNNVSGNDNNV---SGSF---HTVSGHNTVSGSNNTVSGSNNHVVSGT 115
 DB 168 NOAKGEHSTIAGGKNQATGRNSTVAGSNNQAVGTINSTVAGSNNQAKNSFAGVGN 227

RESULT 6
 ID 091961 PRELIMINARY; PRT; 889 AA.
 AC 091961;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE USPA2H.
 GN USPA2H.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OC NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TTA37;
 RX MEDLINE=20138164; PubMed=10671460;
 RA Lafontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr., Hansen E.J.;
 RT "The Uspl1 Protein and a Second Type of Uspl2 Protein Mediate Adherence of Moraxella catarrhalis to Human Epithelial Cells in Vitro."
 RL J. Bacteriol. 182:1364-1373(2000).
 DR EMBL: AF181075; AAF40121.1; -
 SQ SEQUENCE 889 AA; 92184 MW; E82B5BF3C7DE62F1 CRC64;

Query Match 27.2%; Score 166; DB 2; Length 889;
 Best Local Similarity 29.6%; Pred. No. 0.00046;
 Matches 40; Conservative 14; Mismatches 57; Indels 24; Gaps 3;

OY 5 NTIGSNNVTYRSGSKNVLAGNDNTVIGDNNVSGSNNVTYVGSNNHVVSG--T 62
 DB 226 NLAEGKSSAIGGGEFFLALGNNTATISGGRNBSGDRSTVAGGEONQAKSTISGGRQ 285
 OY 63 NHIVTDNNNVSGNDNNVSGSFHTVSGH-----NTVS-GSNNVT 100
 DB 286 NEASGDRSTVAGGEONQAKSTISGGRNBSGDRSTVAGGEONQAKSTISGGRQ 345
 OY 101 VSGSNNHVVSGSKV 115
 DB 346 IEGENSVAGISNNTV 360

RESULT 7
 ID 09XD52 PRELIMINARY; PRT; 912 AA.
 AC 09XD52;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE USPA1.
 GN USPA1.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OC NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V1171;
 RX MEDLINE=99315796; PubMed=10383971;
 RA Cope L.D., Lafontaine E.R., Slaughter C.A., Hasemann C.A. Jr., Aebi C., Henderson F.W., McCracken G.H. Jr., Hansen E.J.;
 RT "Characterization of the Moraxella catarrhalis uspl1 and uspl2 genes and their encoded products."
 RL J. Bacteriol. 181:4026-4034(1999).
 DR EMBL: AF113610; AAD43469.1; -
 DR InterPro: IPR000515; BPD_transp.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.

SEQUENCE 912 AA; 95765 MW; BCB151ADACBCE7EF4 CRC64;

Query Match 25.5%; Score 155.5; DB 2; Length 912;

Best Local Similarity 27.8%; Pred. No. 0.0025;

Matches 37; Conservative 17; Mismatches 44; Indels 35; Gaps 2;

QY 5 NTISGNNNTVRSKKNVLAGNDNTVIGSDNNSVSGSNNNTVSGNDNTVSGNNHYSGTNH 64
 DB 97 NKANGTSTIGGGSYNANKEKSTIGGDDNNTAKGHNSTVGGYKKEATGKYSTVGG--- 153
 QY 65 IVTDNNNNVSGSDNNSVSGSNNNTVSGSNNNTVSGSNNNTVSGSNNNTVSGSNNNTV 102
 DB 154 -----GNSKAKGCTSTIGAGKKNNAKGGSPAGAVEKANKANENAVAGKKNNSIE 203
 QY 103 GSNHNVSGSNKV 115
 DB 204 GKDSVAIGSENTV 216

RESULT 8 24.7%; Score 150.5; DB 2; Length 873;

Best Local Similarity 26.7%; Pred. No. 0.0052;

Matches 40; Conservative 24; Mismatches 47; Indels 39; Gaps 5;

QY 09XD56 PRELIMINARY; PRT; 863 AA.

AC 09XD56;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE USPAL.
 GN USPAL.
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
 OX NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC25238;
 RA MEDLINE=99315796; PubMed=10383971;
 RA Cope L.D., Lafontaine E.R., Slaughter C.A., Hasemann C.A. Jr.,
 RA Aebi C., Henderson F.W., McCracken G.H. Jr., Hansen E.J.;
 RA "Characterization of the Moraxella catarrhalis uspal and uspa2 genes
 RT and their encoded products."
 RL J. Bacteriol. 181:4026-4034(1999).
 DR EMBL; AF113606; AAD43465.1; -
 DR InterPro; IPR000515; BPD.Transp.
 DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
 DR SEQUENCE 863 AA; 90541 MW; FC28EFCBDCFE92 CRC64;

Query Match 25.2%; Score 153.5; DB 2; Length 863;

Best Local Similarity 34.4%; Pred. No. 0.0032;

Matches 43; Conservative 19; Mismatches 44; Indels 19; Gaps 6;

QY 5 NTISGNNNTVRSKKNVLAGNDNTVIGSDNNSVSGSNNNTVSGNDNTVSGN 56
 DB 146 STIGGGGNEATIENTSYGGGYNQAKGRNSTVAGTNNETGTGDSIAGGRKNQAKGK 205
 QY 57 HAVSGTNIHTDNNNNVSGNDNNSVSGSNNNTVSGSNNNTVSGSNNNTVSGSNNK 113
 DB 206 SFAAG-----IDNKANA---DNAVALGNKNTIEGNSVALGSGNNNTVAKGOONFILGNT 257
 QY 114 VTDA 118
 DB 258 DTTNA 262

RESULT 9 23.5%; Score 143.5; DB 5; Length 560;

Best Local Similarity 31.6%; Pred. No. 0.01;

Matches 37; Conservative 25; Mismatches 44; Indels 11; Gaps 3;

QY 09L960 PRELIMINARY; PRT; 873 AA.

AC 09L960;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE USPAL.
 GN USPAL.

OS Moraxella catarrhalis.

OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.

OX NCBI_TaxID=480;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TTA37;

RA MEDLINE=20138164; PubMed=10671460;

RA Lafontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr.,

RA Hansen E.J.;

RA "The uspal Protein and a Second Type of uspa2 Protein Mediate

RT Adherence of Moraxella catarrhalis to Human Epithelial Cells in

RT Viro."

RL J. Bacteriol. 182:1364-1373(2000).

DR EMBL; AF181076; AAF40122.1; -

DR InterPro; IPR000515; BPD.Transp.

DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.

DR SEQUENCE 873 AA; 91880 MW; 175939F71727283C CRC64;

Query Match 24.7%; Score 150.5; DB 2; Length 873;

Best Local Similarity 26.7%; Pred. No. 0.0052;

Matches 40; Conservative 24; Mismatches 47; Indels 39; Gaps 5;

QY 5 NTISGNNNTVRSKKNVLAGNDNTVIGSDNNSVSGSNNNTVSGNDNTVSGNNHYSG--T 62
 DB 114 NEAKGSSSTIGGGDNNNSATGMYSTIGGDDNNSATGRYSTIAGGWLQATGHSSTVAGGWL 173
 QY 63 NHTVDNN-----NNVSGNDNNSVSGSNNNTVSGSNNNTVSGSNNNTVSGSNNNTV 90
 DB 174 NATFNENSTVGGGREFNQAGRNSTVAGGKKNATGVDSITMAGRRNNQANGISFAAGIDN 233
 QY 91 ----HNTVS-GSNNNTVSGSNNHYSGSNKV 115
 DB 234 QANANTVALGNKNIKKKDSVAIGSNNTV 263

RESULT 10

061085 PRELIMINARY; PRT; 560 AA.

AC 061085;

DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)

DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)

DE CORONIN BINDING PROTEIN.

GN DB10.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

OX NCBI_TaxID=4689;

RN [1]

RP SEQUENCE FROM N.A.

RA de Hostos E.L., Benhayon D., Berger M., Gu W.,

RA Subhosted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF051898; AAC05577.1; -

DR SEQUENCE 560 AA; 61281 MW; F417511014157229 CRC64;

Query Match 23.5%; Score 143.5; DB 5; Length 560;

Best Local Similarity 31.6%; Pred. No. 0.01;

Matches 37; Conservative 25; Mismatches 44; Indels 11; Gaps 3;

QY 5 NTISGNNNTVRSKKNVLAGNDNTVIGSDNNSVSGSNNNTVSGNDNTVSGN 63
 DB 389 NNNNNNSNNNSNGNSN--NNNNNNIINNNSNSNSNNNSNNNSNNNSNNNSNNNSNN 446
 QY 64 HAVT-----DNNNVSGNDNNSVSGSNNNTVSGSNNNTVSGSNNNTVSGSNNNTV 112
 DB 447 DNNNTNNTNN 503

RESULT 11

09XYP6 PRELIMINARY; PRT; 718 AA.

AC 09XYP6;

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SNF1/AMP-ACTIVATED KINASE.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.
RN NCBI_TaxID=4689;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3:
RA Sung S., Bissson S., Koehler S., Podgorski G.J.;
RT "The Dictyostellium SNF1/AMP-activated Kinase";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: A118151; AAD30963.2; -
DR HSSP: O63450; 1A06.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_DOM; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 718 AA; 80310 MW; 7F718176EA961C55 CRC64;

Query Match 23.3%; Score 142; DB 5; Length 718;
Best Local Similarity 27.9%; Pred. No. 0.017;
Matches 34; Conservative 27; Mismatches 39; Indels 22; Gaps 3;

OY 5 NTIGSNNNTVRSKVLGNDNTVSGDNNVSQSGN-----NTVSGNDNTVSGSNH 57
DB 394 NSISNNNNNNNNNNNTTNNNNNT-TNNNSTIINNNTNNNNNNNNNNNNNNNNNNI 451
OY 58 VVSGTHIYTDNNNNVSGDNNVSGSFHTVSG-----HNTVSGSNNTVSGS 104
DB 452 INNNTNNNNNNNNNNNNNNNNNNNNSSISGTEVFISPLNNNSYNSGNSGNSGNS 511
OY 105 NH 106
DB 512 NN 513
RESULT 12
OY 098587 PRELIMINARY; PRT; 1176 AA.
AC 098587; 066214;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 119.0 KDA PROTEIN.
GN A540I.
OS Paramoecium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98022962; PubMed=9356347;
RA Li Y., Lu Z., Sun L., Ropp S., Kutish G.F., Rock D.L., Van Etten J.L.;
RT "Analysis of 74 kb of DNA located at the right end of the 330-kb
RT chlorella virus PBCV-1 genome";
RL Virology 237:360-377(1997).
RN [2]
RP SEQUENCE 885-1176 FROM N.A.
RX MEDLINE=9026467; PubMed=2345963;
RA Schuster A.M., Graves M., Korth K., Ziegelbein M., Brumbaugh J.,
RA Grode D., Meints R.H.;
RT "Transcription and sequence studies of a 4.3-kbp fragment from a ds-
RT DNA eukaryotic algal virus";
RL Virology 176:515-523(1990).

DR EMBL: U42580; AAC96907.1; -
DR EMBL: M33758; AAA66404.1; -
DR InterPro: IPR003880; Phosphopant_attach.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Hypothetical protein.
SO SEQUENCE 1176 AA; 118985 MW; 5542053F12AC86ED CRC64;

Query Match 23.1%; Score 141; DB 12; Length 1176;
Best Local Similarity 30.2%; Pred. No. 0.031;
Matches 49; Conservative 27; Mismatches 38; Indels 48; Gaps 10;

OY 5 NTIGS--NNTVRSKSN-----VLGNDNTVIS-----GDNNVSQSG--NNT 43
DB 569 NTIGSYAGNNTTIGSNNNTVIGTFAGTRLIAGNLTIVGFSSGGRGDRNTVGSLSGNTT 628
OY 44 VVSGDNTV-----TGSNHVSGT---NHIVDNNNNVSG-NDNNVS--GSFHTV 87
DB 629 AMTSPDNTIIGSCGPGPLIGTNTVLGSRAGNNLTGSSNTFLGVSTGISTGSLNTA 688
OY 88 SGHNTVSGSNNTVSG--SNHVSG-----SNKVYTD 118
DB 689 VGFSAGTGSSNTVVGAFSGRITGTQNTVFGRAGNLVISA 730

RESULT 13
OY 054356 PRELIMINARY; PRT; 832 AA.
AC 054356;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HIGH MOLECULAR WEIGHT OUTER MEMBRANE PROTEIN.
GN USPAL.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O35E;
RA Hansen E.J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U57551; AAB96359.2; -
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR000719; Euk_pkinase.
DR PROSITE: PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SO SEQUENCE 832 AA; 88292 MW; 5C47F4F27350F4B CRC64;

Query Match 23.0%; Score 140.5; DB 2; Length 832;
Best Local Similarity 27.7%; Pred. No. 0.024;
Matches 31; Conservative 21; Mismatches 59; Indels 1; Gaps 1;

OY 5 NTIGSNNTVRSKSNVLGNDNTVISGDDNNVSQSGNNTVSGDNTVSG-NNHVSQTN 63
DB 111 NKAGRYSTIGGSSNNEATNTETIVGGDDNKATGRYSTIGGGDNNTAEGEYSTVAGGN 170
OY 64 HIVTDNNNNVSGDNNVSGSFHTVSGGNTVSGSNNTVSGSNHVSQSNKV 115
DB 171 NGATGTGSFAGVENVQANAEVAVAKKNIIIECNSVAISSENVTKEHKNV 222

RESULT 14

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li Z., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Styrtas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang M., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003430; AAF45930.1; -
 DR FlyBase: FBgn0029699; CG15575.
 SQ SEQUENCE 337 AA; 36204 MW; F0F7B411AB2F3B5 CRC64;

Query Match 21.8%; Score 133; DB 5; Length 337;
 Best Local Similarity 29.5%; Pred. No. 0.034;
 Matches 31; Conservative 20; Mismatches 50; Indels 4; Gaps 2;

QY 10 SNNVRSKSNVLADNNTVSGDNNVSGSNNNTVSGNDNTVSGSHVYSGNHVYD 69
 DB 223 NNNNTINHSTTF-NNNTTSTNNHSSLDHFTTHNNNTSTNNNTSTNNNTSTN 281
 QY 70 NNNVSGNDNNVSGSFHTVSGGHNVTSGSNNNTVSGSHVYSGSNKV 114
 DB 282 NNNVSTSNNT--YSTTNNNTATATNNCTNDNNHTSFFNNI 323

RESULT 17
 Q90217 PRELIMINARY; PRT; 373 AA.
 ID Q90217
 AC Q90217
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Y11CA4.5 PROTEIN.
 GN Y11CA4.5.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RT none;
 RT "Genome sequence of the nematode *C. elegans*: A platform for
 RT Science 282:2012-2018(1998).
 DR EMBL: AL032627; CAB6354.1; -
 DR InterPro: IPR001079; Gal_bind_Lectin.
 DR SMART: SM00276; GLECT.1.
 SQ SEQUENCE 373 AA; 39451 MW; 577EA334FA30F311 CRC64;

Query Match 21.7%; Score 132.5; DB 5; Length 373;
 Best Local Similarity 36.4%; Pred. No. 0.04;
 Matches 40; Conservative 19; Mismatches 44; Indels 7; Gaps 4;

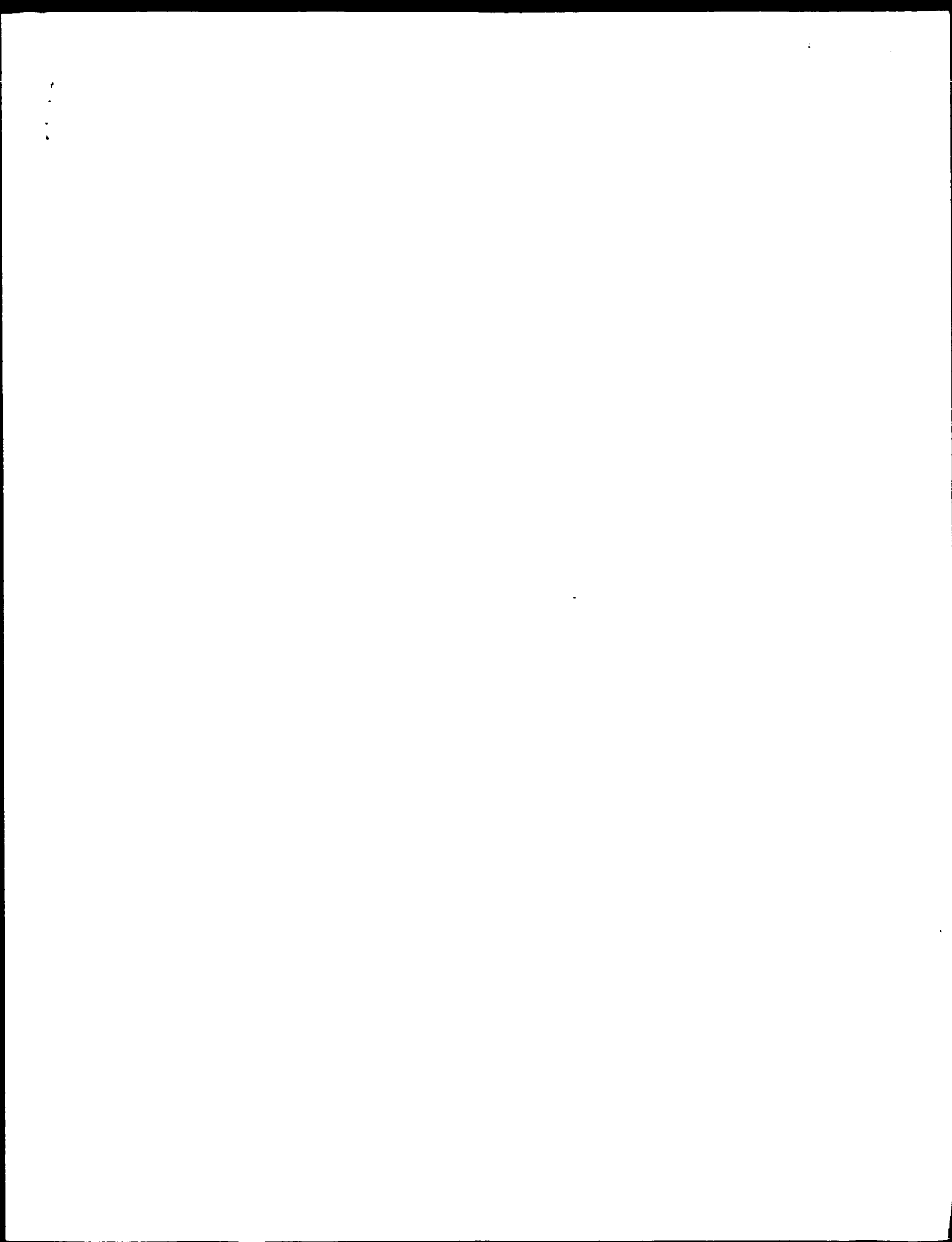
QY 5 NTISSNNVRSKSNVLADNNTVSGDNNVSGSNNNTVSGNDNTVSGSHVYSGNH 64
 DB 175 NTLAEDPNNGQLGGELVYTPNGVSGSGNSNNNNSSNGSN--YGSNNSGSGN 232
 QY 65 IYTDNNNNVSGDNNV--SGSFHTVSGGHNVTSGSNNNTVSGSHVYSGSN 112
 DB 233 SMSGNGNSGNGNMGANGSNGNGSNGNG--NGSNGNGNGNG--NNGN 279

RESULT 18
 Q97257 PRELIMINARY; PRT; 1096 AA.
 ID Q97257
 AC Q97257
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HYPOTHETICAL 134.4 KDA PROTEIN.
 GN PFC0580C, MAL3P5.2.
 OS *Plasmodium falciparum* (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,
 RA Gentles S., Gilliam R., Hain N., Harris D., Holroyd S., Hornsby T.,
 RA Horrocks P., Jagsals K., Jaiswal B., Kyes S., McLean S.,
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skellern J., Squares R., Squares S., Sulston J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrett B.C.,
 RT "The complete nucleotide sequence of chromosome 3 of *Plasmodium*
 RT *falciparum*.";
 RL Nature 400:532-538(1999).
 DR EMBL: AL034556; CAB38968.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1096 AA; 134368 MW; C58757CCACB81A6B CRC64;

Query Match 21.6%; Score 131.5; DB 5; Length 1096;
 Best Local Similarity 27.7%; Pred. No. 0.13;
 Matches 33; Conservative 24; Mismatches 31; Indels 31; Gaps 5;

QY 18 SKNVL-----AGNDNTVSG-----DNNVSGSNNNTVSGNDNTVGSN-HV 59
 DB 555 SKNIYFEDELGRSDNTIYNNVDKETNTTNNNNNDNDNICSNNDHICSNNDHIC 614
 QY 60 SGTN-HIYTDNNNNVSGDNNVSGSFHTVSGGHNVTSGSNNNTVSGSHVYSGSNKVYD 117
 DB 615 SNNNDHICSNNNNNICSNNNN-----NICSNNNNNICSNNNNNICSNNKMLDE 661

RESULT 19
 Q81761 PRELIMINARY; PRT; 343 AA.
 ID Q81761
 AC Q81761
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE HYPOTHETICAL 36.1 KDA PROTEIN.
 GN F1715.120 OR ATAG33930.
 OS *Arabidopsis thaliana* (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2002, 06:44:11 ; Search time 30.24 Seconds

(Without alignments)
433.423 Million cell updates/sec

Title: US-09-600-787-1

Perfect score: 610
Sequence: 1 DEQPMYISGSNNTVRSKSN.....NTVSGSNHVSGSKNKVTDA 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description		
1	610	100.0	118	20	AAV22472	Grass anti-freeze
2	172	28.2	892	19	AAW68203	M. catarrhalis str
3	171	28.0	624	19	AAW68204	M. catarrhalis str
4	169.5	27.8	941	19	AAW68205	M. catarrhalis str
5	166	27.2	889	19	AAW68208	M. catarrhalis str
6	150.5	24.7	873	19	AAW68207	M. catarrhalis str
7	137.5	22.5	831	19	AAW68201	M. catarrhalis str
8	133	21.8	337	22	ABW70781	Drosophila melanog
9	131	21.5	988	22	ABW60007	Drosophila melanog
10	129	21.1	477	22	AAW3085	Amino acid sequenc
11	124.5	20.4	1436	21	AAW18195	Plasmodium falcipa

12	123.5	20.2	1278	21	AAW18277	Plasmodium falcipa
13	123	20.2	342	17	AAW03625	Human lutropin-cho
14	121	19.8	710	22	AAW30814	Amino acid sequenc
15	120	19.7	256	13	AAW24796	Sequence of fibron
16	119	19.5	1997	21	AAW18287	Plasmodium falcipa
17	118	19.3	686	21	AAW18293	Plasmodium falcipa
18	116.5	19.1	190	14	AAW42008	Staphylococcus ent
19	116.5	19.1	666	17	AAW85291	Staphylococcus ent
20	115.5	18.9	590	12	AAW11019	Enterococcus hirae
21	115.5	18.9	964	22	AAW30811	Amino acid sequenc
22	115.5	18.9	2112	22	AAW60403	Drosophila melanog
23	115.5	18.9	2500	21	AAW18272	Plasmodium falcipa
24	115	18.9	718	22	AAW66061	Drosophila melanog
25	113	18.5	190	14	AAW41998	Staphylococcus ent
26	111.5	18.3	190	14	AAW42005	Staphylococcus ent
27	111	18.2	247	22	AAW64742	Staphylococcus ent
28	110.5	18.1	368	22	AAW58030	Drosophila melanog
29	110.5	18.1	645	22	AAW61545	Drosophila melanog
30	110.5	18.1	1601	18	AAW30292	Non-lypeable Haemo
31	110	18.0	190	14	AAW41989	Staphylococcus ent
32	110	18.0	190	14	AAW41990	Staphylococcus ent
33	110	18.0	190	14	AAW41991	Staphylococcus ent
34	110	18.0	190	14	AAW41992	Staphylococcus ent
35	110	18.0	190	14	AAW41993	Staphylococcus ent
36	110	18.0	190	14	AAW41994	Staphylococcus ent
37	110	18.0	190	14	AAW41995	Staphylococcus ent
38	110	18.0	190	14	AAW41996	Staphylococcus ent
39	110	18.0	190	14	AAW41997	Staphylococcus ent
40	110	18.0	190	14	AAW41999	Staphylococcus ent
41	110	18.0	190	14	AAW42000	Staphylococcus ent
42	110	18.0	190	14	AAW42001	Staphylococcus ent
43	110	18.0	190	14	AAW42002	Staphylococcus ent
44	110	18.0	190	14	AAW42003	Staphylococcus ent
45	110	18.0	190	14	AAW42004	Staphylococcus ent
46	110	18.0	190	14	AAW42006	Staphylococcus ent
47	110	18.0	190	14	AAW42007	Staphylococcus ent
48	110	18.0	190	14	AAW42009	Staphylococcus ent
49	110	18.0	190	14	AAW42010	Staphylococcus ent
50	110	18.0	190	14	AAW42011	Staphylococcus ent
51	110	18.0	190	14	AAW42012	Staphylococcus ent
52	110	18.0	190	14	AAW42013	Staphylococcus ent
53	110	18.0	1198	22	AAW60371	Drosophila melanog
54	110	18.0	1222	21	AAW01830	H. influenzae stra
55	110	18.0	1228	21	AAW01828	Haemophilus influe
56	110	18.0	2573	21	AAW18234	Plasmodium falcipa
57	109.5	18.0	319	22	AAW69348	Drosophila melanog
58	109	17.9	1338	14	AAW41731	High molecular wei
59	109	17.9	1529	14	AAW41732	High molecular wei
60	109	17.9	1598	18	AAW30291	Non-lypeable Haemo
61	108.5	17.8	180	22	AAW31927	Peptide #4578 enco
62	108.5	17.8	180	22	AAW37161	Peptide #4667 enco
63	108.5	17.8	180	22	AAW22471	Protein #4470 enco
64	108.5	17.8	180	22	AAW57877	Human brain expres
65	108.5	17.8	180	22	AAW70299	Human bone marrow
66	108.5	17.8	180	22	AAW18131	Peptide #4565 enco
67	108.5	17.8	180	22	AAW30636	Peptide #4673 enco
68	108.5	17.8	180	22	AAW05762	Peptide #4444 enco
69	108.5	17.8	365	22	AAW64606	H. pylori HPC137 p
70	107.5	17.6	719	22	AAW70700	S cerevisiae apopt
71	107.5	17.6	1250	22	AAW12254	Human S3-12 homolo
72	106.5	17.5	550	22	AAW70830	C albicans apoptos
73	106.5	17.5	1909	22	AAW24938	Novel human diagno
74	106	17.4	668	18	AAW55483	H. pylori ORF 14ap
75	106	17.4	668	22	AAW64377	H. pylori HFN137 p
76	106	17.4	677	18	AAW55328	H. pylori ORF hp3e
77	106	17.4	871	21	AAW95550	Chlamydia pneumoni
78	106	17.4	1368	22	AAW60262	Drosophila melanog
79	105.5	17.3	657	22	AAW57783	Drosophila melanog
80	105.5	17.3	992	21	AAW01843	Haemophilus influe
81	105.5	17.3	998	21	AAW01842	Haemophilus influe
82	105	17.2	969	21	AAW01827	Haemophilus influe
83	105	17.2	975	21	AAW01826	Haemophilus influe
84	105	17.2	1221	21	AAW01825	Haemophilus influe

```

85 105 17.2 1227 21 AAB01824 Haemophilus influenzae
86 104.5 17.1 1778 22 ABB52677 Escherichia coli p
87 104.5 17.1 1992 17 AAM04505 Moraxella 200 kDa
88 104.5 17.1 1992 22 AAB69133 M. catarrhalis str
89 104.5 17.1 1992 22 AAB69133 M. catarrhalis M56
90 104.5 17.1 2047 22 AAB69134 M. catarrhalis str
91 104.5 17.1 2053 22 AAB69135 M. catarrhalis str
92 104.5 17.1 2123 22 AAE00701 Moraxella catarrha
93 104 17.0 2367 22 AAB60025 Drosophila melanog
94 104 17.0 1073 21 AAB01837 Novel human secret
95 103.5 17.0 1079 21 AAB01836 Haemophilus influenzae
96 103.5 17.0 1079 21 AAB01836 Haemophilus influenzae
97 103.5 17.0 2314 22 AAB69136 M. catarrhalis les
98 103 16.9 117 22 ABB28811 Peptide #1462 enco
99 103 16.9 117 22 ABB33996 Peptide #1502 enco
100 103 16.9 117 22 ABB19436 Protein #1435 enco

```

ALIGNMENTS

RESULT 1

AAV22472 standard; Protein; 118 AA.

AAV22472;

29-SEP-1999 (first entry)

Grass anti-freeze protein sequence.

Anti-freeze protein; grass; plant; frozen food product; frost tolerance; frozen confectionery.

Loium perenne.

WO937782-A2.

29-JUL-1999.

23-DEC-1998; 98WO-EP08553.

22-JAN-1998; 98GB-0001408.

(UNIL) UNILEVER NV.

(UNIL) UNILEVER PLC.

Jarman CD, Sidebottom CM, Twigg S, Morrall D;

WPI; 1999-458697/38.

N-PSDB; AAX99717.

New plant anti-freeze protein useful in frozen food products

Claim 3; Page 36; 39pp; English.

This sequence is the plant anti-freeze protein of the invention. The anti-freeze protein is characterised in that at least 40% of its amino acids are from the group of serine, threonine and asparagine. The anti-freeze protein can be used in frozen food products, especially frozen confectionery. Anti-freeze proteins are especially used in food products, which are heated, e.g. by pasteurisation, blanching or sterilisation prior to freezing. Plants transformed with a nucleic acid sequence encoding the anti-freeze protein have an increased frost tolerance. Prior art anti-freeze proteins have not been applied to commercially available food products, due to high costs and complicated processes for obtaining the protein. Also prior art anti-freeze proteins have tended to destabilise during processing especially during the pasteurisation step. This is overcome by the present anti-freeze protein. The anti-freeze proteins provide an ice particle size following an ice recrystallisation inhibition assay of 15 µm or less. The anti-freeze protein ingredient means that mixes can be frozen under quiescent conditions, e.g. in a shop or home freezer without the

CC formation of unacceptable ice crystal shapes and hence with a texture
CC different to products normally obtained via quiescent freezing.

XX Sequence 118 AA;

Query Match 100.0%; Score 610; DB 20; Length 118;

Best Local Similarity 100.0%; Pred. No. 6,5e-47; Mismatches 0; Indels 0; Gaps 0;

Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DEOPTSGSNNTVRSQSKVTLAGNDNTVISCNNNSVSGSNNTVSGNDNTVSGSNHVS 60

1 deopntsgsnntvrsqskvltlagndntviscnnnsvsghntvsgndntvsgsnhvs 60

61 GTNHIVTDNNNNVSGNDNNVSGSFHTVSGGHNTVSGSNNTVSGSNHVS 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

61 gtnhivtdnnnnvsgndnnvsgsfhtvsgghntvsgsnntvsgsnhvs 118

Query Match 28.2%; Score 172; DB 19; Length 892;
Best Local Similarity 32.7%; Pred. No. 3.8e-07;
Matches 35; Conservative 19; Mismatches 47; Indels 6; Gaps 1

OY 9 GSNNTVRSGSKNVLANDNDTVITISGDNNSVSGSNNFTVSGNDNTVTGSNHVVSGTNHIVTD 68
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 143 gdsstisgygynqatgbekstvaggrnmdqtgnmtsvagsgyngdltmtnstvaggsh---- 198

OY 69 NNNNVSGDNNNSVGSPFHTVSGGHNTVSGSNNFTVSSGNHHVSGSKKV 115
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 199 --ngdtggsfaagvenkanannavaljgnntldgdnsvalgsnltl 243

RESULT 3
AAW68204
ID AAW68204 standard; Protein; 624 AA.
XX AC AAW68204;
XX DT 07-OCT-1998 (first entry)
DE M. catarrhalis strain O46E UspA2 antigen.
XX Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
XX immunity enhancer; immunoassay reagent.
OS Moraxella catarrhalis.
PN W09828333-A2.
PD PD
XX 02-JUL-1998.
PF 19-DEC-1997; 97WO-USJ3930.
PR 20-DEC-1996; 96US-0033598.
PA (TEXA) UNIV TEXAS SYSTEM.
XX Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;
PI Mactaver I;
XX WP1; 1998-377595/32.
DR N-P5DB; AAV41344.
XX New peptide(s) containing the core epitope of Moraxella catarrhalis
PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.
PT catarrhalis infection, and antibodies for passive immunisation
PS Claim 28; Pages 147-149; 237pp; English.

This represents a UspA2 antigen of Moraxella catarrhalis strain O46E.
Nucleic acid sequences encoding the UspA1 and A2 antigens of
M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
genetic vaccination. An antigenic composition or vaccine containing
antigenic peptides from UspA1 or UspA2 antigens are used to induce an
immune response in mammals against M. catarrhalis and can be used to
treat infections such as otitis media, sinusitis, lower respiratory
tract infections. They can also be used as immunity enhancers for other
bacterial, parasitic or viral antigens, to raise antibodies and as
immunoassay reagents for detecting specific antibodies. The antibodies
are useful for passive immunisation and as immunoassay reagents.
Detection of the epitopic core sequence, by immunoassay or by PCR, is
used to diagnose infection. The usp antigens encoding nucleic acid
sequences are also used to produce recombinant proteins and for screening
for potential anti-M. catarrhalis agents, while their fragments are
useful as diagnostic probes or primers or to isolate variant sequences.

Berry Match 28.0%; Score 171; DB 19; Length 624;
Query Similarity 34.7%; Pred. No. 3,1e-07;
Best Local Alignment

Matches 43; Conservative 19; Mismatches 50; Indels 12; Gaps 4

OY 5 NTISGSNNVTSGSKVLACGNDNTVTSIGNNSVSQSNNTTVYSGNDNVTYTGSHNVSGT-N 63
Db |::| :||| || ::||| :||| :||| :||
104 ngakghstltagsgnqatgrntstvagsnmhvgtlnltvlgssunqdkansfaagavn 163
OY 64 HIVDNL-----NKNVGSDNNVNVSSEFHTVSGGHMTV--SCSNNTVSGSNHWVG---SN 112
Db ||| ||::||| ||::||| | ||| :||| :||| :||
164 qantcnavaajgknmlngmnsaaigseentvmengkvnflgsnttnagsgsvllghetsg 223
OY 113 KVRT 116
Db | |
224 Keat 227

RESULT 4
ID AAM68205
AA AAM68205 standard; Protein; 941 AA.
AC AAM68205;
DT 07-OCT-1998 (first entry)
DE M. catarrhalis strain TTA24 UspA1 antigen.
XX Moraxella catarrhalis; Uspa1; Uspa2; antigen; genetic vaccination;
KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
KM immunity enhancer; immunoassay reagent.
XX Moraxella catarrhalis.
OS WO9828333-A2.
PN 02-JUL-1998.
PD 19-DEC-1997; 97MO-US23930.
PF 20-DEC-1996; 96US-0033598.
PR (TEXA) UNIT TEXAS SYSTEM.
PA Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ,
PI MacIver I;
PT WPI; 1998-377595/32.
DR N-PDSB; AAV41345.
XX New peptide(s) containing the core epitope of Moraxella catarrhalis
PT USP proteins - useful in, e.g. vaccines to prevent or treat M.
PT catarrhalis infection, and antibodies for passive immunisation
PS Claim 30; Pages 152-155; 237pp; English.

This represents a Uspl1 antigen of Moraxella catarrhalis strain TTA24. Nucleic acid sequences encoding the USPl1 and A2 antigens of M. catarrhalis isolates OJ5E, O46E, TTA24 and TTA37 can be used in genetic vaccination. An antigenic composition or vaccine containing antigenic peptides from USpl1 or USpa2 antigens are used to induce an immune response in mammals against M. catarrhalis and can be used to treat infections such as otitis media, sinusitis, lower respiratory tract infections. They can also be used as immunity enhancers for other bacterial, parasitic or viral antigens, to raise antibodies and as immunoassay reagents for detecting specific antibodies. The antibodies are useful for passive immunisation and as immunoassay reagents. Detection of the epitopic core sequence, by immunoassay or by PCR, is used to diagnose infection. The usp antigens encoding nucleic acid sequences are also used to produce recombinant proteins and for screening for potential anti-M. catarrhalis agents, while their fragments are useful as diagnostic probes or primers or to isolate variant sequences.

CC This represents a *UspspA* antigen of *Moraxella catarrhalis* strain TPA37.
CC Nucleic acid sequences encoding the *Uspsp1* and *A2* antigens of
CC *M. catarrhalis* isolates O35F, O46E, TPA24 and TPA37 can be used in
CC genetic vaccination. An antigenic composition or vaccine containing
CC antigenic peptides from *Uspsp1* or *Uspsp2* antigens are used to induce an
CC immune response in mammals against *M. catarrhalis* and can be used to
CC treat infections such as otitis media, sinusitis, lower respiratory
CC tract infections. They can also be used as immunoenhancers for other
CC bacterial, parasitic or viral antigens, to raise antibodies and as
CC immunoassay reagents for detecting specific antibodies. The antibodies
CC are useful for passive immunisation and as immunoassay reagents.
CC Detection of the epitopic core sequence, by immunoassay or by PCR, is
CC used to diagnose infection. The *Usp* antigens encoding nucleic acid
CC sequences are also used to produce recombinant proteins and for screening
CC for potential anti-*M. catarrhalis* agents, while their fragments are
CC useful as diagnostic probes or primers or to isolate variant sequences.

PS Chalm 34, Pages 161-164; 237pp. English.

XX

CC This represents a Usral antigen of *Moraxella catarrhalis* strain TTA37.

CC Nucleic acid sequences encoding the Usral and A2 antigens of

CC *M. catarrhalis* isolates 0359, 046E, TTA21 and TTA37 can be used in

CC genetic vaccination. An antigenic composition or vaccine containing

CC antigenic peptides from Usral or Uspl2 antigens are used to induce an

CC immune response in mammals against *M. catarrhalis* and can be used to

CC treat infections such as otitis media, sinusitis, lower respiratory

CC tract infections. They can also be used as immunity enhancers for other

CC bacterial, parasitic or viral antigens, to raise antibodies and as

CC immunosay reagents for detecting specific antibodies. The antibodies

CC are useful for passive immunisation and as immunosay reagents.

CC detection of the epitopic core sequence, by immunosay or by PCR, is

CC used to diagnose infection. The uspl antigens encoding nucleic acid

CC sequences are also used to produce recombinant proteins and for screening

CC for potential anti-*M. catarrhalis* agents, while their fragments are

CC useful as diagnostic probes or primers or to isolate variant sequences.

XX	Sequence	873 AA:
XX	Query Match	24.7%; Score 150.5; DB 19; Length 873;
XX	Best Local Similarity	26.7%; Pred. No. 2.9e-05;
XX	Matches 40; Conservative 24; Mismatches 47; Indels 39; Gaps 5	
OY	5 NTIGSGNNTRSSSKNLVAGNDQTVTSGDNNVSGSNNNTVYSGNDQNTVGTSGNHVSG--T	62
DB	114 neakgesstllyggdnnmsatgymstlyggdnnmsatgrystlaagwlnqatglsstvagw	173
OY	63 NHIYTDNN-----NNVSGDNNVSGSF-----HTVSGG-----	90
DB	174 nqatnenstvyggrfrngatgrmstvagykknkatgvydstlaagrmnganglgsstaagldn	233
OY	91 ----HMTVS-GSNNTVYSGSNHVYSGSNKVV	115
DB	234 qanantvtvalgnknlkgykdsvalgsanltv	263
RESULT	7	
AA668201	standard; Protein; 831 AA.	
AA668201;		
07-OCT-1998	(first entry)	
M. catarrhalis strain O35E	Usplal antigen.	
Moraxella catarrhalis: Usplal; UsplA2; antigen; genetic vaccination;		
vaccline; otitis media; sinusitis; lower respiratory tract infection;		
immunity enhancer; immunoassay reagent.		
Moraxella catarrhalis.		
MO9828333-A2.		
02-JUL-1998.		
19-DEC-1997;	97MO-US23930.	
20-DEC-1996;	96US-0033598.	
(TEXA) UNIV TEXAS SYSTEM.		
Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen ET;		
Maciver I;		
WPI; 1998-377595/32.		
N-PSDB; AAV41341.		
New peptide(s) containing the core epitope of Moraxella catarrhalis		
Usp proteins - useful in, e.g. vaccines to prevent or treat M.		
catarrhalis infection, and antibodies for passive immunisation		
Claim 22; Pages 133-136; 237pp; English.		
This represents a Usplal antigen of Moraxella catarrhalis strain O35E.		
Nucleic acid sequences encoding the Usplal and A2 antigens of		
M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in		
genetic vaccination. An antigenic composition or vaccine containing		
antigenic peptides from Usplal or UsplA2 antigens are used to induce an		
immune response in mammals against M. catarrhalis and can be used to		
treat infections such as otitis media, sinusitis, lower respiratory		
tract infections. They can also be used as immunity enhancers for other		
bacterial, parasitic or viral antigens, to raise antibodies and as		
immunoassay reagents for detecting specific antibodies. The antibodies		
are useful for passive immunisation and as immunoassay reagents.		
Detection of the epitopic core sequence, by immunoassay or by PCR, is		
used to diagnose infection. The usp antigens encoding nucleic acid		
sequences are also used to produce recombinant proteins and for screening		

Query Match	22.5%;	Score 137.5;	DB 19;	Length 831;
Best Local Similarity	27.7%;	Pred. No. 0.00039;		
Matches 31;	Conservative 21;	Mismatches 59;	Indels 1;	Gaps 1
OY	5	NTIGSNNTVYRSGSKNVLAGNDNTVYISGDNNSVSGSNNTVYSGNDNTVYTG-	NHNVSGTN 63	
Db	110	nkaegrystllygssnneatneytlvgddnkatgrystllygddnhtregeystvaggn	169	
OY	64	HTVDDNNNNVSGNDNNVSGSFHTVSGGHTVYSGSNNTVYSGNHVYSGSNKYV	115	
Db	170	ngatcgtstfaaygenganaenavaygkniiiegensvaigseantvktekhv	221	
RESULT	8			
ID	ABB70781	standard; Protein; 337 AA.		
XX	ABB70781;			
XX	26-MAR-2002	(first entry)		
XX	Drosophila melanogaster	polypeptide SEQ ID NO 39135.		
XX	Drosophila;	developmental biology; cell signalling; insecticide;		
XX	pharmaceutical.			
XX	Drosophila melanogaster.			
XX	WO200171042-A2.			
XX	27-SEP-2001.			
XX	23-MAR-2001;	2001WO-US09231.		
XX	23-MAR-2000;	2000US-191637P.		
XX	11-JUL-2000;	2000US-0614150.		
XX	(PEKE)	PE CORP NY.		
XX	Venter JC, Adams M, Li PWD, Myers EW;			
XX	WPI; 2001-656860/75.			
XX	N-PSDB; ABL14884.			
XX	New isolated nucleic acid detection reagent for detecting 1000 or more			
XX	genes from Drosophila and for elucidating cell signalling and cell-cell			
XX	interactions -			
XX	Disclosure; SEQ ID NO 39135; 21pp + Sequence Listing; English.			
XX	The invention relates to an isolated nucleic acid detection reagent			
XX	capable of detecting 1000 or more genes from Drosophila. The invention is			
XX	useful in developmental biology and in elucidating cell signalling and			
XX	cell-cell interactions in higher eukaryotes for the development of			
XX	insecticides, therapeutics and pharmaceutical drugs. The invention			
XX	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
XX	sequences (ABL01840-ABL16175) and the encoded proteins			
XX	(ABB57737-ABB72072).			
XX	The sequence data for this patent did not form part of the printed			
XX	specification, but was obtained in electronic format directly from WIPO			
XX	at ftp.wipo.int/pub/published_pat_sequences.			
XX	Sequence 337 AA;			
XX	Query Match	21.8%;	Score 133;	DB 22;
XX	Best Local Similarity	29.5%;	Pred No. 0.00034;	Length 337;

Query Match	18.98;	Score 115.5;	DB 22;	Length 964
Best Local Similarity	32.68;	Pred. No. 0.041;		

QY 5 NTLSGSNNNTVBSGSKNVLAQDNNTVLSODNNVSGSNNTVSGNDNTVSGSHHVGSTNH 64
| | | | | : | | : : : : : | | : | | | | | | | |
Db 332 nqsgsngnrgs -sgsagsnsnsgnsnngnqsgnsgnsngsssssq -sgnsq -ssgsnq 386
QY 65 IYVDNNNNNTVSGSFHYVSGSHHTVSGSNNTVSGSNHVVSGSKNV 115
: | | : | | | | : : : | | | | | | | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2002, 06:44:41 ; Search time 13.26 Seconds

(without alignments)
217.362 Million cell updates/sec

Title: US-09-600-787-1

Perfect score: 610
Sequence: 1 DEQPTIGSNNTYRSGSKN.....NTVSGSNHYVSGSKVVTDA 118Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	172	28.2	892	4	US-09-336-447A-5
2	171	28.0	624	4	US-09-336-447A-7
3	169.5	27.8	941	4	US-09-336-447A-9
4	166	27.2	889	4	US-09-336-447A-15
5	150.5	24.7	873	4	US-09-336-447A-13
6	137.5	22.5	831	4	US-09-336-447A-1
7	123	20.2	110	2	US-08-569-166-34
8	116.5	19.1	666	2	US-08-737-716-14
9	109	17.9	1338	2	US-08-728-470-9
10	109	17.9	1338	2	US-08-719-641-9
11	109	17.9	1529	2	US-08-728-470-10
12	109	17.9	1529	2	US-08-719-641-10
13	109	17.9	1599	2	US-08-617-697-9
14	109	17.9	1600	2	US-08-617-697-10
15	104.5	17.1	2123	4	US-08-968-685A-10
16	103.5	16.9	1536	1	US-09-268-347-49
17	103	16.9	1536	1	US-08-038-682-2
18	103	16.9	1536	1	US-08-302-832-2
19	103	16.9	1536	2	US-08-530-198-2
20	103	16.9	1536	2	US-08-469-880-2
21	103	16.9	1536	2	US-08-728-470-2
22	103	16.9	1536	2	US-08-617-697-2
23	103	16.9	1536	2	US-08-719-641-2
24	102.5	16.8	2048	4	US-09-268-347-48
25	101.5	16.6	246	4	US-09-451-117-2
26	99.5	16.3	1612	1	US-08-169-927-2
27	95.5	15.7	361	1	US-08-415-751-3

28	93.5	15.3	379	1	US-07-723-002C-6	Sequence 6, App1
29	93	15.2	1477	1	US-08-038-682-4	Sequence 4, App1
30	93	15.2	1477	1	US-08-302-832-4	Sequence 4, App1
31	93	15.2	1477	2	US-08-530-198-4	Sequence 4, App1
32	93	15.2	1477	2	US-08-469-880-4	Sequence 4, App1
33	93	15.2	1477	2	US-08-728-470-4	Sequence 4, App1
34	93	15.2	1477	2	US-08-617-697-4	Sequence 4, App1
35	93	15.2	1477	2	US-08-719-641-4	Sequence 4, App1
36	92.5	15.2	430	2	US-08-945-848-8	Sequence 8, App1
37	92.5	15.2	670	2	US-08-737-716-13	Sequence 6, App1
38	90.5	14.8	888	2	US-08-861-464-6	Sequence 6, App1
39	90.5	14.8	888	2	US-08-336-001-6	Sequence 6, App1
40	90.5	14.8	888	2	US-09-323-433A-6	Sequence 6, App1
41	90	14.8	137	2	US-08-630-822A-64	Sequence 64, App1
42	90	14.8	137	2	US-09-005-059-64	Sequence 2, App1
43	90	14.8	2391	2	US-08-446-855A-2	Sequence 2, App1
44	90	14.8	2391	4	US-09-150-741-2	Sequence 113, App
45	88	14.4	630	1	US-08-487-890A-113	Sequence 113, App
46	88	14.4	630	2	US-08-478-435-113	Sequence 113, App
47	88	14.4	630	2	US-08-337-483-113	Sequence 113, App
48	88	14.4	630	3	US-08-478-373-113	Sequence 113, App
49	88	14.4	630	3	US-08-474-671-113	Sequence 113, App
50	88	14.4	630	3	US-08-483-577A-113	Sequence 113, App
51	88	14.4	630	4	US-08-897-438-113	Sequence 113, App
52	88	14.4	630	4	US-08-637-654-113	Sequence 113, App
53	87.5	14.3	654	1	US-08-487-850A-12	Sequence 12, App1
54	87.5	14.3	654	2	US-08-478-435-12	Sequence 12, App1
55	87.5	14.3	654	2	US-08-337-483-12	Sequence 12, App1
56	87.5	14.3	654	2	US-08-478-373-12	Sequence 12, App1
57	87.5	14.3	654	3	US-08-474-671-12	Sequence 12, App1
58	87.5	14.3	654	3	US-08-483-577A-12	Sequence 12, App1
59	87.5	14.3	654	4	US-08-897-438-12	Sequence 12, App1
60	87.5	14.3	654	4	US-08-637-654-12	Sequence 12, App1
61	87	14.3	236	4	US-09-605-858-34	Sequence 34, App1
62	86.5	14.2	1004	4	US-09-268-347-30	Sequence 30, App1
63	86	14.1	1094	4	US-09-268-347-32	Sequence 32, App1
64	85	13.9	1912	1	US-08-409-995-4	Sequence 4, App1
65	85	13.9	1912	3	US-08-685-467-4	Sequence 4, App1
66	85	13.9	2353	4	US-09-377-155-33	Sequence 33, App1
67	85	13.9	2353	4	US-08-913-942-4	Sequence 4, App1
68	85	13.9	2353	4	US-09-669-974-33	Sequence 33, App1
69	85	13.9	2354	4	US-09-268-347-47	Sequence 47, App1
70	85	13.9	2411	4	US-09-268-347-36	Sequence 36, App1
71	85	13.9	3060	2	US-08-487-826B-14	Sequence 14, App1
72	84	13.8	681	4	US-08-760-615-6	Sequence 6, App1
73	83.5	13.7	461	1	US-08-186-222-2	Sequence 2, App1
74	83	13.6	631	1	US-08-487-890A-111	Sequence 111, App
75	83	13.6	631	2	US-08-478-435-111	Sequence 111, App
76	83	13.6	631	2	US-08-337-483-111	Sequence 111, App
77	83	13.6	631	2	US-08-478-373-111	Sequence 111, App
78	83	13.6	631	3	US-08-474-671-111	Sequence 111, App
79	83	13.6	631	3	US-08-483-577A-111	Sequence 111, App
80	83	13.6	631	4	US-08-897-438-111	Sequence 111, App
81	83	13.6	631	4	US-08-637-654-111	Sequence 111, App
82	83	13.6	937	3	US-09-005-180A-4	Sequence 4, App1
83	83	13.6	1003	1	US-08-571-758-4	Sequence 4, App1
84	83	13.6	1003	1	US-08-909-984A-4	Sequence 4, App1
85	83	13.6	1003	1	US-08-909-983-4	Sequence 4, App1
86	82.5	13.5	1026	1	US-08-194-290-7	Sequence 7, App1
87	82.5	13.5	1026	2	US-08-614-377A-7	Sequence 7, App1
88	82.5	13.5	1026	2	US-09-142-648B-7	Sequence 7, App1
89	81	13.3	385	2	US-08-387-942C-8	Sequence 8, App1
90	81	13.3	970	1	US-08-375-709-7	Sequence 7, App1
91	81	13.3	970	1	US-08-752-929-7	Sequence 7, App1
92	81	13.3	970	4	US-09-090-793-5	Sequence 5, App1
93	80.5	13.2	376	1	US-08-594-031-100	Sequence 100, App
94	80.5	13.2	376	1	US-08-594-031-102	Sequence 102, App
95	80.5	13.2	408	2	US-09-118-319-5	Sequence 5, App1
96	80.5	13.2	508	2	US-08-818-024-3	Sequence 3, App1
97	80.5	13.2	508	4	US-09-334-775A-3	Sequence 3, App1
98	80	13.1	1186	1	US-08-485-568A-4	Sequence 4, App1
99	80	13.1	1186	1	US-08-357-699-6	Sequence 6, App1
100	80	13.1	1186	2	US-08-590-554A-4	Sequence 4, App1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 17.9%; Score 109; DB 4; Length 1529;
Best Local Similarity 26.8%; Pred. No. 0.026;
Matches 30; Conservative 24; Mismatches 46; Indels 12; Gaps 5;

QY 8 SGSNNTVRSKSNVLAGNDNTVISCDDNSVSG---SNNTVSGNDNTVSGSNHVSCTNH 64
DB 1072 AAGNVTYTKEGT-TINATGSGVEVTAQNGITKGNITSONVTYATENLVTTENAVINATSG 1130
QY 65 IYTDNNNNVSGNDNNVSGFHTVSGGHNTVSGSNNTVSGSNHVSCKNYVT 116
DB 1131 TV-----NISTKTGDIKGIESTSGNVN-ITASGNTLKYSN--ITGDVYTVT 1174

RESULT 13
US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9

Query Match 17.9%; Score 109; DB 2; Length 1599;
Best Local Similarity 26.8%; Pred. No. 0.027;
Matches 30; Conservative 24; Mismatches 46; Indels 12; Gaps 5;

QY 8 SGSNNTVRSKSNVLAGNDNTVISCDDNSVSG---SNNTVSGNDNTVSGSNHVSCTNH 64
DB 1141 AAGNVTYTKEGT-TINATGSGVEVTAQNGITKGNITSONVTYATENLVTTENAVINATSG 1199
QY 65 IYTDNNNNVSGNDNNVSGFHTVSGGHNTVSGSNNTVSGSNHVSCKNYVT 116
DB 1200 TV-----NISTKTGDIKGIESTSGNVN-ITASGNTLKYSN--ITGDVYTVT 1243

RESULT 14
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 17.9%; Score 109; DB 2; Length 1600;
Best Local Similarity 26.8%; Pred. No. 0.027;
Matches 30; Conservative 24; Mismatches 46; Indels 12; Gaps 5;

QY 8 SGNVNVSSGSKNVLGNDNTVIGSDNNVSG---SNNTVVSGNDNTVGTGSHVSGENH 64

Db 1142 AAGVTTKEGT-TTNATGTSVEVTAQNGTICKNTSSONTVATATENLVTEMNAVINTSG 1200

QY 65 IVTDNNNNVSGDNNVSGSFHTVSGGHNTVSGSNNTVSGSNHVVSGSNKYVT 116

Db 1201 TV-----NISRTGDKIGKESTSGSNVN-ITASGNTLTKVSN--ITGDDVTVT 1244

RESULT 15
US-08-968-685A-10

; Sequence 10, Application US/08968685A

; Patent No. 6214981

; GENERAL INFORMATION:

; APPLICANT: TUCKER, KENNETH

; APPLICANT: PLOSTLA, LAURA

; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: No. 6214981ember 12, 1997

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Baldwin, Geraldine F.

; REGISTRATION NUMBER: 31,232

; REFERENCE/DOCKET NUMBER: 7969-060

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864

; TELEEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2123 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

; US-08-968-685A-10

Query Match 17.1%; Score 104.5; DB 4; Length 2123;
Best Local Similarity 27.4%; Pred. No. 0.1;
Matches 29; Conservative 21; Mismatches 43; Indels 13; Gaps 4;

QY 15 RSGSKNVLGNDNTVIGSDNNVSGSNNTVS-----GNDNTVGTGSHVSGTGNHIVTD 68

Db 1864 QAGNOSTAIG-DNAQATGDSIAIGTGNVYAGKHSAGIADPSTVKADNSYSVGNNGFTD 1922

QY 69 -NNNNVSGDNNVSGSFHTVSGGHNTVSGSNNTVSGSNHVVSGSNK 113

Db 1923 ATQTDVFGVGNNT-----TVTESNSVALGSNSAISAGTHAGTQAKK 1963

RESULT 16

US-09-268-347-49

; Sequence 49, Application US/09268347

; Patent No. 6335182

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

; FILE REFERENCE: 1038-860

; CURRENT APPLICATION NUMBER: US/09/268,347

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 49

; LENGTH: 2314

; TYPE: PRP

; ORGANISM: Haemophilus influenzae

Query Match 17.0%; Score 103.5; DB 4; Length 2314;

Best Local Similarity 31.3%; Pred. No. 0.14; Mismatches 45; Indels 5; Gaps 5;

Matches 31; Conservative 18; Mismatches 45; Indels 5; Gaps 5;

QY 23 AGNDNTVIGSDNNVSGSNNTVVSGNDNTVGTGSHVSGTGNHIVTDNNNNVSGDNNV 81

Db 2056 AGNQAIAI-DNAQAATG-DQSAIGTGNVYVTKHSGAIGDPSTVKADNSYSVGNNGFTD 2113

QY 82 GSFHTVSG-GHN-TVSGSNNTVSGSNHVVSGSNKYVTDA 118

Db 2114 ATQTDVFGVGNNTVTVESNSVALGSNSAISAGTHAGTQAKK 2152

RESULT 17

US-08-038-682-2

; Sequence 2, Application US/08038682

; Patent No. 5549897

; GENERAL INFORMATION:

; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: SR. GEME III, JOSEPH W

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 16-MAR-1993

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: BERKSTRESSER, JERRY W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-293

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1536 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-038-682-2

```

Query Match	16.9%;	Score 103;	DB 1;	Length 1536;
Best Local Similarity	25.2%;	Pred. No. 0.095;		
Matches 35;	Conservative 19;	Mismatches 45;	Indels 40;	Gaps 6;

[illegible]

RESULT 18
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938

```

1  GENERAL INFORMATION:
2  APPLICANT: Barenkamp, Stephen J
3  TITLE OF INVENTION: High Molecular Weight Surface Proteins
4  TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
5  NUMBER OF SEQUENCES: 8
6  CORRESPONDENCE ADDRESS:
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
10
```

ADDRESS: SNOOMAKER AND MATTHEW, LTD.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:

Query Match	16.9%;	Score 103;	DB 1;	Length 1536;
Best Local Similarity	25.2%;	Pred. No. 0.095;		

Matches 35; Conservative 19; Mismatches 45; Indels 40; Gaps 6;

QY	6	TISSNNSTVSGSKNNVLAIGDNTVIGSDNNVSGSNNTVSGNDNTVYGSNNVSGTNHI	65
		: : : : : : : : : : :	
Db	1277	TSISNTVYVATNAGDVLVGN-----CAEINATEGAAATLTSSGKLTTEASHITSAKQV	1331
QY	66	VTDDNNNNVSGNDNNVSGSFH-----TVSGGHTVYSGSN-NTVSGS-----	104
		: : : : : : : :	
Db	1332	-----NLSAQDGSVAGSITAAAVTLNTGTTLTYVGSNNINATSGTIVLNKADLNLGAA	1385
QY	105	--NHVV-----SGSKRV	115
		: : :	
Db	1386	LGHTTVYVATNANGSGSVI	1404

RESULT 19
US-08-530-198-2
; Sequence 2, Application US/08530198
; Patent No. 5869065
GENERAL INFORMATION:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1 ADDRESSSEE: Shoemaker and Mattare, Ltd
2 STREET: 2001 Jefferson Davis Hwy., 1203 Crysta
3 STREET: Bldg. 1
4 CITY: Arlington
5 STATE: Virginia
6 COUNTRY: U.S.A.
7 ZIP: 22202-0286
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: PatentIn Release #1.0, Version #1.25
14
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/530,198

Query Match	16.9%;	Score 103;	DB 2;	Length 1536;
Best Local Similarly	25.2%;	Pred. No. 0.095;		
Matches 35;	Conservative 19;	Mismatches 45;	Indels 40;	Gaps 6

QY	6	T T S S N N T V S G S K N N T A G D N T V I S G D N N V S G S N N T V A S G D N D T V T S G S N H I	65
Db	1277	T I S G T V A N V T A N A G D L V G N - - - - G A E I N A T E G A T L T T S S G K L T T E A S S H I T S A K O V	1333
		: : : :	
QY	66	V T D N N N V S G D N N V S G S F H - - - - T V S G H N T V S G S N - N T V S G S - - - - -	104
		: : : : : : : :	
Db	1332	- - - - - N I S A Q D G S V A S I N A A N V T L N T G T G L T T V A S G S N I N A T S G T I V I N A K D A E L N G A A	1385
		: : : : : : : :	
QY	105	- - - - - N H V - - - - - S G S K N V	115
		: :	

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

Query Match 16.9%; Score 103; DB 2; Length 1536;
Best Local Similarity 25.2%; Pred. No. 0.095;
Matches 35; Conservative 19; Mismatches 45; Indels 40; Gaps 6;
QY 6 TISGNNVTVRSGSKNYLAGNDNTVISGDNNSVSGSNNNTVSGNDNTVGSNNHVSQTNHI 65
DB 1277 TISGNNVTVRSGSKNYLAGNDNTVISGDNNSVSGSNNNTVSGNDNTVGSNNHVSQTNHI 1331
QY 66 VTDDNNNNVSGNDNNVSGSFH-----TVSGGHNVTVSGSN-NTVSGS----- 104
DB 1332 -----NLSSADGSAVAGSINAAVTLNTGTTLTVKGSNNINATSGTLVINAKDAELNGAA 1385
QY 105 --NHVY-----SGSNKYV 115
DB 1386 LGNHTVVAATNANGSGSVI 1404

RESULT 23
US-08-719-641-2
Sequence 2, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-719-641-2

Query Match 16.9%; Score 103; DB 4; Length 1536;
Best Local Similarity 25.2%; Pred. No. 0.095;
Matches 35; Conservative 19; Mismatches 45; Indels 40; Gaps 6;
QY 6 TISGNNVTVRSGSKNYLAGNDNTVISGDNNSVSGSNNNTVSGNDNTVGSNNHVSQTNHI 65
DB 1277 TISGNNVTVRSGSKNYLAGNDNTVISGDNNSVSGSNNNTVSGNDNTVGSNNHVSQTNHI 1331
QY 66 VTDDNNNNVSGNDNNVSGSFH-----TVSGGHNVTVSGSN-NTVSGS----- 104
DB 1332 -----NLSSADGSAVAGSINAAVTLNTGTTLTVKGSNNINATSGTLVINAKDAELNGAA 1385
QY 105 --NHVY-----SGSNKYV 115
DB 1386 LGNHTVVAATNANGSGSVI 1404

RESULT 24
US-09-268-347-48
Sequence 48, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 48
LENGTH: 2048
TYPE: PRT
ORGANISM: Haemophilus influenzae

US-09-268-347-48

Query Match

16.8%; Score 102.5; DB 4; Length 2048;

Best Local Similarity 27.4%; Pred. No. 0.15; Mismatches 44; Indels 13; Gaps 4;

Matches 29; Conservative 20; Mismatches 44; Indels 13; Gaps 4;

QY 15 RSGSKNVLAGNDNTVYISGDNNSVSGSNTTVS-----GNDNTVYSGSNHVVSGTNHIVTD 68

DB 1789 QAGNQSIALIG-DNAQATGDSIALGRTNVVAGKHGSAIGDPSYKADNSYSGNNQFTD 1847

QY 69 -NNNNVSGDNNVSGSFHTVSGHNNTVYSGSNHVVSGSNK 113

DB 1848 ATQTDVFGVGNNI-----TYTENSVALGSNSAISAGTHAGTQAKK 1888

RESULT 25

US-09-451-117-2

; Sequence 2, Application US/09451117

; Patent No. 6277973

; GENERAL INFORMATION:

; APPLICANT: Jenkins, Mark C.

; APPLICANT: Fayer, Ronald

; APPLICANT: Trout, James

; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41

; Patent No. 6277973

; TITLE OF INVENTION: Kda Cryptosporidium parvum Oocyst Wall Protein

; FILE REFERENCE: 0046, 99

; CURRENT APPLICATION NUMBER: US/09/451,117

; CURRENT FILING DATE: 1999-11-30

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patentln Ver. 2.1

; SEQ ID NO 2

; LENGTH: 246

; TYPE: PRT

; ORGANISM: Cryptosporidium parvum

US-09-451-117-2

Query Match

16.6%; Score 101.5; DB 4; Length 246;

Best Local Similarity 27.6%; Pred. No. 0.014;

Matches 32; Conservative 29; Mismatches 48; Indels 7; Gaps 4;

QY 2 EOPNTISGSNTVRSKSNVLAGNDNTVYISGDNNSVS--GSNNTVSG--NDNTVTGSNH 57

DB 129 EYTDNIGMGIITSSGDSIAVYTN--LNGNNSNSISGNFIYPVGTCSSTISGNSG 186

QY 58 VVSGTNHIVTDNNNNVSGDNNVSG-SFHTVSGHNNTVYSGSNHVVSGSN 112

DB 187 VAFTAIHNNNNNSNNNNNNNNNSNTLTVAATNANITTTTNTTTTNNNNNN 242

Search completed: July 29, 2002, 06:49:44
Job time: 303 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2002, 06:46:11 ; Search time 16.69 seconds

(Without alignments)
679.361 Million cell updates/sec

Title: US-09-600-787-1

Perfect score: 610
Sequence: 1 DEQPMTIGSGNNVTFRSGSKN.....NTVSGSNHVSQSNKVYVDA 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	173.5	28.4	622	2 AF0169	probable exported
2	149	24.4	420	2 AE1857	hypothetical prote
3	142	23.3	448	2 S05355	hypothetical prote
4	141	23.1	1176	2 T18042	ice nucleation pro
5	139.5	22.9	1650	2 T18444	hypothetical prote
6	136	22.3	255	2 AF1857	hypothetical prote
7	135	22.1	1584	2 T18276	protein-tyrosine k
8	134.5	22.0	1275	2 T33369	hypothetical prote
9	133	21.8	286	2 C61615	sericin MG-2 - gre
10	131	21.5	343	2 T05221	hypothetical prote
11	131	21.5	974	1 A40213	optic lobe develop
12	124.5	20.4	1436	2 D71618	hypothetical prote
13	123.5	20.2	1256	2 S14556	asparagine-rich pr
14	123.5	20.2	1278	2 A71609	probable secreted
15	122.5	20.1	2150	2 S71629	sensory transducti
16	120	19.7	234	2 S14469	asparagine-rich pr
17	120	19.7	309	2 AD0144	hypothetical prote
18	119	19.5	967	2 S6852	DNA helicase II BR
19	119	19.5	1997	2 F71607	asparagine-rich pr
20	118.5	19.4	239	2 S14470	cell surface glyco
21	118.5	19.4	1797	2 F69195	Mtn3/RAGLR-like p
22	118	19.3	686	2 A71607	nuclear pore compl
23	117.5	19.3	959	2 B44402	hypothetical prote
24	117	19.2	721	2 S29795	lysosome 2 (EC 3.2
25	116.5	19.1	666	2 A44296	probable membrane
26	115.5	18.9	964	2 S48404	asparagine/asparta
27	115.5	18.9	1360	2 T18403	hypothetical prote
28	115.5	18.9	2500	2 G71609	probable surface p
29	115	18.9	658	2 AH0110	

30	115	18.9	718	1 S33168	gene pointed prote
31	114.5	18.8	451	2 A23535	clustered asparagi
32	114	18.7	709	2 T28712	hypothetical prote
33	114	18.7	720	2 T51007	hypothetical prote
34	114	18.7	1004	2 C82672	surface-exposed ou
35	113.5	18.6	1093	2 T18275	1-phosphatidylinos
36	113.5	18.6	1649	2 C86822	hypothetical prote
37	113	18.5	490	2 S52830	HMS1 protein - yea
38	112.5	18.4	430	1 HHBYD8	heat shock protein
39	112	18.4	782	2 A10062	conserved hypotnet
40	110.5	18.1	568	2 UC7210	molluscan shell ma
41	110.5	18.1	642	2 S27806	homeotic protein B
42	110.5	18.1	937	2 S58135	hypally regulated
43	110	18.0	669	2 S14535	asparagine-rich pr
44	110	18.0	799	2 B66735	hypothetical prote
45	110	18.0	900	2 AC1852	hypothetical prote
46	110	18.0	2573	2 D71614	hypothetical prote
47	109.5	18.0	313	2 T04776	hypothetical prote
48	109.5	18.0	355	2 C39725	hypothetical prote
49	109.5	18.0	1374	2 AE3259	extracellular seri
50	109.5	18.0	1487	2 AG2560	hypothetical prote
51	108.5	17.8	173	2 A47303	FTF-P1 steroid rec
52	108.5	17.8	640	2 A41726	homeotic protein B
53	108.5	17.8	1585	2 T18274	1-phosphatidylinos
54	108.5	17.8	2708	2 T09079	probable chloroqui
55	108	17.7	431	2 S50977	hypothetical prote
56	108	17.7	537	2 A23770	asparagine-rich pr
57	108	17.7	3848	2 T17414	FTIC protein - ali
58	108	17.7	4936	2 AH2515	hypothetical prote
59	107.5	17.6	719	2 S61046	ARPI protein - yea
60	107.5	17.6	783	2 E97340	uncharacterized co
61	107.5	17.6	1102	2 A32247	virg protein - Shi
62	107	17.5	1314	1 TBNBYR6	transcription regu
63	107	17.5	1317	2 A54831	nuclear pore compl
64	106.5	17.5	1367	2 T33819	hypothetical prote
65	106	17.4	398	2 H70393	hemolysin - Aquife
66	106	17.4	623	1 S33167	gene pointed prote
67	106	17.4	630	2 S28796	hypothetical prote
68	106	17.4	668	2 C71868	hypothetical prote
69	106	17.4	927	2 AG1739	transmembrane prot
70	106	17.4	1407	2 B72078	polymorphic outer
71	105.5	17.3	543	2 S19933	glycine-rich prote
72	105.5	17.3	552	2 D70604	probable PPE prote
73	105.5	17.3	676	2 S61977	transcription fact
74	105.5	17.3	2020	2 C48399	ABC-type transport
75	105	17.2	419	2 T18420	hypothetical prote
76	105	17.2	795	2 A69196	cell surface glyco
77	105	17.2	795	2 AE2444	hypothetical prote
78	104.5	17.1	368	2 H86990	probable membrane
79	104.5	17.1	698	2 A69222	hypothetical prote
80	104.5	17.1	1119	2 T11491	hypothetical prote
81	104	17.0	364	2 AE0169	probable exported
82	104	17.0	836	2 T18460	hypothetical prote
83	104	17.0	4550	2 T18440	hypothetical prote
84	103.5	17.0	666	2 S50452	hypothetical prote
85	103.5	17.0	741	2 AC0094	conserved hypotnet
86	103.5	17.0	1225	2 T09057	probable protein-h
87	103.5	17.0	1274	2 S59405	probable membrane
88	103.5	17.0	1341	2 S50366	probable membrane
89	103	16.9	419	2 S14508	asparagine-rich pr
90	103	16.9	894	2 S51245	probable finger pr
91	103	16.9	943	2 S68824	cyto
92	103	16.9	1536	2 A43855	high-molecular-we
93	103	16.9	1844	2 D71612	hypothetical prote
94	103	16.9	1858	2 T18273	1-phosphatidylinos
95	103	16.9	3097	2 T28635	glutamate synthase
96	102.5	16.8	507	2 A71622	hypothetical prote
97	102.5	16.8	614	2 F66719	hypothetical prote
98	102.5	16.8	1034	2 JC2143	ice nucleation act
99	102.5	16.8	1200	1 SNPSO	ice nucleation pro
100	102.5	16.8	1258	2 JQ0188	ice nucleation pro

Db 629 AMTSEDTIIGSCGPIIGTNTLVGSRAGNNLTGSGNFFLGSTGISTGSLNTA 688
 QY 88 SGGHNTVSGSNTVSG--SNHVSG-----SNKVYVDA 118
 Db 689 VGFSAAGTSSNTVVGAFSGERITOTNTVFVGFAGNIVTSA 730

RESULT 5
 18444
 hypothetical protein C0385C - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18444
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18444
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1650 <LAW>
 A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB1112.2
 C:Genetics:
 A:Map position: 3
 A:Introns: 1597/3; 1625/3
 A:Note: C0385C

Query Match 22.9%; Score 139.5; DB 2; Length 1650;
 Best Local Similarity 28.0%; Pred. No. 0.0026;
 Matches 33; Conservative 36; Mismatches 18; Indels 31; Gaps 8;

QY 1 DEQPTIGSNNNTVSGSKNYLAGNDNTVIGSDNNVSGSNNNTVSGNDNTVGSNNHVS 60
 Db 126 DSSNNMNGNNN-----MNGNNN--MNGNNMNGNNN--MNGNNN--MNGNNN--MN 169
 QY 61 GTNHVTVNNNNVSGNDNNVSGSFHTVSGNHTVSGSN-----NTVSGSNNHVSNNKV 114
 Db 170 GNNNN--MDNNNNMNGNN-----IMNGNNIMNGNNNNINNNIMNGNNIMNGNNNI 216

RESULT 6
 AF1857
 hypothetical protein all0407 [imported] - Anabaena sp. (strain PCC 7120)
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AF1857
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF1857
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-235 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA072365.1; PID:g17129752; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0407

Query Match 22.3%; Score 136; DB 2; Length 255;
 Best Local Similarity 28.4%; Pred. No. 0.00062;
 Matches 31; Conservative 18; Mismatches 58; Indels 2; Gaps 2;

QY 5 NTIGSNNNTVSGSKNYLAGNDNTVIGSDNNVSGSNNNTVSGNDNTVGSNNHVSSTNH 64
 Db 138 NPFAGSNPFAGGNNPFAGGSNP-LTGGGNPFAGGSNPFAGGNNPFAGGSNPFAGGSN 196
 QY 65 IVTNNNNVSGNDNNVSGSFHTVSGGHT--VSGSNNNTVSGSNNHVSSTNH 112
 Db 197 PFAGSNPFAGGSNPFAGGSNPFAGGSNPFAGGSNPFAGGSNPFAGGSNPFAGGSN 245

RESULT 7
 T18276
 protein-tyrosine kinase (EC 2.7.1.112) 1 - slime mold (Dictyostellium discoideum)
 C:Species: Dictyostellium discoideum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
 C:Accession: T18276; A35670
 R:Nickolls, G.H.; Oshero, N.; Loomis, W.F.; Spudich, J.A.

Development 122, 3295-3305, 1996
 A:Title: The Dictyostellium dual-specificity kinase SPLA is essential for spore differ
 A:Reference number: Z18852; MUID:97053827
 A:Accession: T18276
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1564 <NUC>
 A:Cross-references: EMBL:U32174; NID:g974333; PID:g974334; PIDN:AAB41125.1
 A:Tan, J.L.; Spudich, J.A.
 Mol. Cell. Biol. 10, 3578-3583, 1990
 A:Title: Developmentally regulated protein-tyrosine kinase genes in Dictyostellium dis
 A:Reference number: A35670; MUID:90287147

A:Accession: A35670
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'R', 1249-1434, 'L', 1436-1584 <TRAN>
 A:Cross-references: GB:M33785; NID:g167775; PIDN:AAA33202.1; PID:g167776
 A:Note: the authors translated the codon TAA for residue 271519 as Thr
 C:Genetics:
 A:Gene: SPLA
 A:Introns: 47/3; 72/2
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
 FJ1287-1566/Domain: protein kinase homology <KIN>
 FJ1295-1303/Region: protein kinase ATP-binding motif

Query Match 22.1%; Score 135; DB 2; Length 1584;
 Best Local Similarity 35.5%; Pred. No. 0.0056;
 Matches 39; Conservative 18; Mismatches 33; Indels 20; Gaps 5;

QY 5 NTIGSNNNTVSGSKNYLAGNDNTVIGSDNNVSGSNNNTVSGNDNTVGSNNHVSSTNH 64
 Db 468 NNNNNNNNNNSNSST--NNNNNTTNNNNNSNN--NNNNNSNSNN--NNNNNSNN 519
 QY 65 IVTNNNNVSGNDNNV-----SGSFHTVSGGHTVSGSNNNTVSGS 104
 Db 520 I--NNNNNNNNNNNTVLRKPSIGSTDESSSTGSLGNNSSGNNSSSSS 567

RESULT 8
 T33369
 hypothetical protein H02F09.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T33369
 R:Geisler, C.; Harmon, G.
 submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid H02F09.
 A:Reference number: Z21330
 A:Accession: T33369
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1275 <GEI>
 A:Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
 A:Experimental source: strain Bristol N2; clone H02F09
 C:Genetics:
 A:Gene: CESP:H02F09.3
 A:Map position: X
 A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
 C:Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 22.0%; Score 134.5; DB 2; Length 1275;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2002, 06:49:26 ; Search time 11.89 seconds

(without alignments)
384.265 Million cell updates/sec

Title: US-09-600-787-1

Perfect score: 610

Sequence: 1 DEQPTIGSGNNVTYRSGSKN.....NTVSGSNHVVSGSKVYTD 118

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt-40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	23.3	448	1 AACQ_DICDI	P14196 dictyostell
2	135	22.1	1584	1 KYKI_DICDI	P18160 dictyostell
3	131	21.5	988	1 OMB_DROME	Q24432 drosophila
4	118.5	19.4	793	1 REGA_DICDI	Q23917 dictyostell
5	117.5	19.3	959	1 N100_YEAST	Q02629 saccharomyc
6	117	19.2	721	1 YCEZ_OENPI	P31568 oenothera p
7	116.5	19.1	666	1 MUR2_ENTHR	P39046 enterococcu
8	116	19.0	989	1 PTR3_DICDI	P54637 dictyostell
9	115.5	18.9	964	1 Y1NO_YEAST	P40467 saccharomyc
10	115	18.9	718	1 PNT2_DROME	P51023 drosophila
11	115	18.9	1953	1 BIGA_SALTY	P25927 salmoneilla
12	114.5	18.8	451	1 ARP2_PLAFA	P13824 plasmodium
13	114	18.7	1093	1 PI4K_DICDI	P54785 dictyostell
14	113	18.5	490	1 MOT3_YEAST	P54785 dictyostell
15	112.5	18.4	429	1 DR48_YEAST	P18899 saccharomyc
16	110.5	18.1	937	1 HYR1_CANAL	P46591 candida alb
17	108.5	17.8	1585	1 P3K2_DICDI	P54675 dictyostell
18	108	17.7	537	1 ARP_PLAFA	P04931 plasmodium
19	107.5	17.6	719	1 NRPI_YEAST	P32770 saccharomyc
20	107.5	17.6	1196	1 ICEY_PSEEX	O33479 pseudomonas
21	107	17.5	1314	1 SW11_YEAST	P09547 saccharomyc
22	107	17.5	1317	1 N145_YEAST	P49687 saccharomyc
23	106	17.4	630	1 PNT1_DROME	P51022 drosophila
24	106	17.4	630	1 YCF2_OENVI	P31569 oenothera v
25	105.5	17.3	373	1 PGLR_GIBPU	O07181 gibberella
26	105.5	17.3	676	1 RLM1_YEAST	Q12224 saccharomyc
27	105.5	17.3	2003	1 YDBA_ECOLI	P33666 escherichia
28	104.5	17.1	860	1 VG12_BBP03	Q37893 bacterioph
29	103.5	17.0	451	1 MYB8_DICDI	P34127 dictyostell
30	103.5	17.0	666	1 YEAT_YEAST	P40002 saccharomyc
31	103.5	17.0	1341	1 YL78_YEAST	O05854 saccharomyc
32	103	16.9	1858	1 P3K2_DICDI	P54674 dictyostell
33	102.5	16.8	1034	1 ICEN_PANAN	Q47879 pantoea ana

34	102.5	16.8	1200	1 ICEN_PSEEX	P06620 pseudomonas
35	102.5	16.8	1258	1 ICEN_ERHME	P16239 erwinia her
36	102.5	16.8	1322	1 ICEN_PANAN	P20469 pantoea ana
37	102	16.7	1365	1 SUZ2_DROME	P25172 drosophila
38	101.5	16.6	472	1 NU49_YEAST	Q02199 saccharomyc
39	101.5	16.6	1076	1 NUP1_YEAST	P20676 saccharomyc
40	101.5	16.6	1656	1 OMBP_RICIA	O06653 r outer mem
41	101	16.6	220	1 SER2_GALME	O06615 galliera me
42	100.5	16.5	1015	1 PMPH_CHLTR	O04880 chlamydia t
43	100	16.4	1655	1 OMBP_RICCN	O06653 r outer mem
44	99.5	16.3	1569	1 YPJA_ECOLI	P21443 escherichia
45	99.5	16.3	1643	1 OMBP_RICPR	O03020 r outer mem
46	99	16.2	429	1 HML2_HELBO	O01622 helobdella
47	99	16.2	1286	1 AIDA_ECOLI	O03155 escherichia
48	99	16.2	1286	1 AIDA_ECOLI	O03155 escherichia
49	98.5	16.1	176	1 V1PA_MYCR	P27625 plasmodium
50	98.5	16.1	648	1 KAPC_DICDI	P29228 mycoplasma
51	98.5	16.1	1094	1 YB00_YEAST	P34099 dictyostell
52	98.5	16.1	1148	1 ICEN_PSEEX	P38114 saccharomyc
53	98	16.1	3726	1 TRX_DROME	O30611 pseudomonas
54	97	15.9	928	1 PM10_CHLPN	P20659 drosophila
55	96.5	15.8	400	1 PMOA_DICDI	O9rb65 chlamydia p
56	96.5	15.8	1567	1 ICEN_XANCT	P34681 dictyostell
57	96.5	15.8	2175	1 HMCU_DROME	P18127 xanthomonas
58	95.5	15.7	1113	1 N116_YEAST	P10180 drosophila
59	95.5	15.7	1250	1 YPAL_ECOLI	Q02630 saccharomyc
60	94.5	15.5	823	1 NSPI_YEAST	P45508 escherichia
61	94.5	15.5	1210	1 ICEN_PSEPL	P14907 saccharomyc
62	94.5	15.5	2329	1 YS89_CAEEL	P09815 pseudomonas
63	94	15.4	671	1 HMOQ_DROME	O09624 caenorhadi
64	93.5	15.3	389	1 PLYB_ASFPG	P22810 drosophila
65	93.5	15.3	378	1 SERI_BOMMO	O00205 aspergillus
66	93.5	15.3	721	1 OGP_MOUSE	P00856 bombyx mori
67	93	15.2	914	1 AZFL_YEAST	O62010 mus musculu
68	93	15.2	1026	1 STRA_DROME	P41696 saccharomyc
69	92.5	15.1	671	1 ALYS_ENTFA	P25159 drosophila
70	92	15.1	1570	1 P3K1_DICDI	P37710 enterococcu
71	92	15.1	2249	1 OMPA_RICRI	P54673 dictyostell
72	91.5	15.0	534	1 GCR2_YEAST	P15921 rickettsia
73	91.5	15.0	1080	1 NRK1_YEAST	O01722 saccharomyc
74	91	14.9	823	1 SCH9_YEAST	P38692 saccharomyc
75	91	14.9	1036	1 HP12_DEIRA	P11792 saccharomyc
76	91	14.9	1150	1 APNU_PIG	P13126 deinococcus
77	91	14.9	1407	1 CYAA_DICDI	P12021 sus scrofa
78	91	14.9	1419	1 MDR_PLAFA	O03100 dictyostell
79	91	14.9	1596	1 MAM_DROME	P13568 plasmodium
80	90.5	14.8	526	1 ZABA_YEAST	P21519 drosophila
81	90.5	14.8	742	1 KM65_YEAST	O00362 saccharomyc
82	90.5	14.8	888	1 YB84_YEAST	Q03656 saccharomyc
83	90.5	14.8	952	1 PM16_CHLPN	P25338 saccharomyc
84	90.5	14.8	1577	1 HYVA_PROMI	O92882 chlamydia p
85	90	14.8	952	1 VO14_BACAN	P16466 proteus mir
86	89.5	14.7	380	1 PIXA_COGL	O9rn18 bacillus an
87	89.5	14.7	731	1 BAF1_YEAST	O00374 collettotric
88	89.5	14.7	858	1 CYAG_DICDI	P14164 saccharomyc
89	89.5	14.7	1276	1 PMP6_CHLPN	Q03101 dictyostell
90	89.5	14.7	2452	1 RB01_PLAFA	O92899 chlamydia p
91	89	14.6	575	1 VG05_BPT4	P14248 plasmodium
92	88.5	14.5	491	1 ZAPA_PROMI	P16009 bacterioph
93	88.5	14.5	854	1 VG12_BPH2	O11137 proteus mir
94	88.5	14.5	854	1 VG12_BPH2	P20346 bacterioph
95	88.5	14.5	970	1 Y277_MYCGE	P07537 bacterioph
96	88	14.4	344	1 ULB63_MYCVA	Q49409 mycoplasma
97	88	14.4	1300	1 I20K_RICRI	P16833 human cytom
98	88	14.4	1654	1 OMBP_RICRI	P14914 rickettsia
99	87.5	14.3	788	1 SPL_RAT	O53047 r outer mem
100	87.5	14.3	1905	1 TAGB_DICDI	O01714 rictus norv
					P24683 dictyostell

ALIGNMENTS

RESULT 1

CC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.
CC	NCBI_TaxId=44689;
CC	11]
CC	SEQUENCE FROM N.A.
CC	STRATIN-JH10;
CC	MedLINE=97053827; PubMed=8898241;
CC	Nickolls G.H., Oshero N., Loomis W.F., Spudich J.A.;
CC	"The Dictyostellum dual-specificity kinase sp1a is essential for
CC	spore differentiation.";
CC	Development 122:3295-3305(1996).
CC	[2]
CC	SEQUENCE OF 1248-1584 FROM N.A.
CC	MedLINE=90287147; PubMed=1972546;
CC	Tan J.L., Spudich J.A.;
CC	"Developmentally regulated protein-tyrosine kinase genes in
CC	dictyostellum discoideum.";
CC	Mol. Cell. Biol. 10:3578-3583(1990).
CC	-1- FUNCTION: ESSENTIAL FOR SPORE DIFFERENTIATION.
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC	tyrosine phosphate.
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH A PEAK
CC	DURING THE MOUND STAGE OF MORPHOGENESIS.
CC	-1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES BUT ALSO TO
CC	SERINE/THREONINE PROTEIN KINASES.
CC	-1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its use
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U32174; AAB41125.1; -
DR	EMBL; M33785; AAA33202.1; -
DR	PIR; A35670; A35670.
DR	HSSP; P11362; IAGW.
DR	Dictpdb; DP03010; pykA.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR001660; SAM.
DR	InterPro; IPR003877; SPRY.
DR	InterPro; IPR003878; SPRY_domain.
DR	InterPro; IPR004040; STY_pkinase.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00356; SAM; 1.
DR	Pfam; PF00622; SPRY; 3.
DR	SMART; SM00454; SAM; 1.
DR	SMART; SM00449; SPRY; 3.
DR	SMART; SM00221; STYKc; 1.
DR	PROSITE; PSS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PSS0109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PSS0105; SAM_DOMAIN; 1.
KW	transferrase; tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT	DOMAIN 908 972 SAM.
FT	DOMAIN 403 420 POLY-ASN.
FT	DOMAIN 428 435 POLY-THR.
FT	DOMAIN 449 480 POLY-ASN.
FT	DOMAIN 483 491 POLY-ASN.
FT	DOMAIN 494 508 POLY-ASN.
FT	DOMAIN 512 532 POLY-ASN.
FT	DOMAIN 596 600 POLY-ASN.
FT	DOMAIN 808 811 POLY-PHE.
FT	DOMAIN 1026 1029 POLY-SER.
FT	DOMAIN 1195 1210 POLY-ASN.
FT	DOMAIN 1215 1220 POLY-GLN.
FT	DOMAIN 1224 1233 POLY-GLN.
FT	DOMAIN 1266 1274 POLY-PRO.
FT	DOMAIN 1289 1561 PROTEIN_KINASE.
FT	NP_BIND 1295 1303 ATP (BY SIMILARITY).
FT	BINDING 1316 1316 ATP (BY SIMILARITY).

FT ACT_SITE 1417 1417 BY SIMILARITY.
FT CONFLICT 1248 1248 D -> R (IN REF. 2).
FT CONFLICT 1435 1435 V -> L (IN REF. 2).
SQ SEQUENCE 1584 AA; 174304 MW; 5D1589458DB01E3 CRC64;
Query Match 22.18; Score 135; DB 1; Length 1584;
Best Local Similarity 35.5%; Pred. No. 0.0025;
Matches 39; Conservative 18; Mismatches 33; Indels 20; Gaps 5;
QY 5 NTGISNNNTVRSKNTLAGNDNTVYISGDNNSVSGSNNTVSGNDNTVYSGNHTVSGTIN 64
D 468 NNNNNNNNNNNNNSSNT---NNNNNTTNNNNNSNNN---NNNNNSNSN---SNNNN 519
QY 65 IVTDNNNNVSGNDNNV-----SGSPFTVSGCHNTVSGSNNTVSGS 104
D 520 I---NN 567
RESULT 3
OMB DROME STANDARD; PRT; 988 AA.
ID OMB DROME 024432; 027917; 09W4K5;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Optomotor-blind protein (lethal(1)optomotor-blind) (L(1)omb) (Blf1d protein).
GN Bl OR OMB OR CG3578.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pelegronota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN-Oregon-R; TISSUE-Embryo;
MEDLINE=92159016; PubMed=1741374;
RA Pflugfelder G.O., Roth H., Poock B., Kerschner S., Schwarz H.,
RA Jonschker B., Helsenberg M.,
RT "The lethal(1)optomotor-blind gene of Drosophila melanogaster is a
RT major organizer of optic lobe development: isolation and
RT characterization of the gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:1199-1203(1992).
[2]
SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bhatnagar S.,
RA Botchko D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fjosek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberg C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Raibeck K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[3]
SEQUENCE OF 1-447 FROM N.A., AND MUTATIONAL ANALYSIS.
RP TISSUE=Larva;
RC MEDLINE=93261414; PubMed=8492800;
RX Poock B., Ballew J., Pflugfelder G.O.;
RT "Transcript identification in the optomotor-blind locus of Drosophila
RT melanogaster by intragenic recombination mapping and PCR-aided
RT sequence analysis of lethal point mutations.";
RL Mol. Gen. Genet. 238:325-332(1993).
CC -1- FUNCTION: ESSENTIAL PROTEIN THAT MAY FUNCTION AS A TRANSCRIPTION
CC REGULATOR. FLIES WITH L(1)OMB MUTATIONS SHOW SEVERE MALDEVELOPMENT
CC OF THE OPTIC LOBES, REDUCTION IN WING SIZE AND AN INCREASED
CC ABDOMINAL PIGMENTATION. THEY DIE DURING THE PUPAL STAGE.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: IN THIRD-INSTAR LARVAE IT IS FOUND IN THE
CC BRAIN REGION THAT WILL DEVELOP INTO OPTIC LOBES AND MORE WEAKLY IN
CC THE THORACIC PART OF THE VENTRAL GANGLION.
CC -1- DEVELOPMENTAL STAGE: THE PEAK PERIODS OF EXPRESSION ARE: MID-
CC EMBRYOGENESIS, THE SECOND DAY OF PUPAL DEVELOPMENT AND IN THE
CC ADULT.
CC -1- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL, M81796; AAA28736.1; -;
DR EMBL, AE003431; AAF45946.1; -;
DR EMBL, S61732; AAB26697.1; -;
DR EMBL, S61727; AAB26697.1; JOINED.
DR EMBL, S61729; AAB26697.1; JOINED.
DR EMBL, S61744; AAB26699.1; -;
DR EMBL, S61743; AAB26699.1; JOINED.
DR HSSP; P24781; 1XBR.
DR Flybase; FBgn0000179; bl.
DR InterPro; IPR001699; T-box.
DR Pfam; PF00907; T-box; 1.
DR PRINTS; PR00937; TBOX.
DR SMART; SM00425; TBOX; 1.
DR PROSITE; PS01283; TBOX_1; 1.
DR PROSITE; PS01264; TBOX_2; 1.
DR PROSITE; PS00552; TBOX_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 51 140 ASN-RICH.
FT DOMAIN 104 107 POLY-THR.
FT DOMAIN 179 184 POLY-SER.
FT DOMAIN 229 236 POLY-GLN.
FT DOMAIN 338 244 POLY-PRO.
FT DNA-BIND 332 513 T-BOX.
FT DOMAIN 574 577 POLY-ASP.
FT DOMAIN 607 692 ALA-RICH.
FT DOMAIN 823 831 POLY-GLY.
FT DOMAIN 910 916 POLY-ALA.
FT DOMAIN 926 966 GLN/HIS-RICH.
FT CONFLICT 10 10 F -> L (IN REF. 1 AND 3).

FT CONFLICT 216 216 A -> P (IN REF. 1 AND 3).
 FT CONFLICT 511 511 F -> L (IN REF. 1).
 FT CONFLICT 823 823 MISSING (IN REF. 1).
 FT CONFLICT 976 988 MISSING (IN REF. 1).
 SQ SEQUENCE 988 AA; 103992 MW; 032B7A4471743FC9 CRC64;

Query Match 21.5%; Score 131; DB 1; Length 988;
 Best Local Similarity 34.1%; Pred. No. 0.003;
 Matches 42; Conservative 17; Mismatches 42; Indels 22; Gaps 7;

QY 2 EQPRTTSGNNVTSGSKNVLGNDNTVIGSDNNVSGNNVTGSDNTV-----TG-- 54
 DB 35 QRPDEFSV-SLTAGSNNNSGNTN--SGNNNSNNNTNTNTNTNULVAVSPTGGG 90
 QY 55 -----SNHYVSGTNIHTVNTNNNNVSGNDNNVSGSFHTVSGH--NTVSGSNNVTGSGNH 106
 DB 91 AQLSPQSNH--SSSNTTTSNTNTNNSSSNNNN--NNSTHNNNNHTNNNNNNNNNTSQKQGH 147
 QY 107 VVS 109
 DB 148 HLS 150

RESULT 4
 REGA.DICDI STANDARD; PRT; 793 AA.
 ID REGA.DICDI
 AC 023917;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 3',5'-cyclic-nucleotide phosphodiesterase REGA (EC 3.1.4.17) (PDBase
 DE REGA).
 GN Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_Taxid=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA MEDLINE=97140317; PubMed=8986798;
 RA Shaulsky G., Escalante R., Loomis W.F.;
 RT "Developmental signal transduction pathways uncovered by genetic
 RT suppressors";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX2;
 RA Thomson P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
 RA Kay R.R.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MORPHOLOGICAL SUPPRESSOR OF TAGB.
 CC -1- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
 CC nucleoside 5'-phosphate.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS IN VEGETATIVE CELLS
 CC AND AT HIGH LEVELS IN PRESPORE AND PRESTALK CELLS DURING
 CC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U60170; AAB03508.1;
 CC EMBL: AJ005398; CAA06513.1;
 CC DR Dictydb; DD02055; rega.
 DR InterPro; IPR003607; HDC.

DR InterPro; IPR002073; PDase.
 DR InterPro; IPR001789; Response-reg.
 DR Pfam; PF00273; PDase; 1.
 DR Pfam; PF00272; response-reg; 1.
 DR SMART; SM00471; HDC; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS00126; PDASE_1; 1.
 DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
 KW Hydrolase; CAMP; cAMP.
 FT DOMAIN 161 280
 FT DOMAIN 18 28
 FT DOMAIN 52 69
 FT DOMAIN 87 96
 FT DOMAIN 100 121
 FT DOMAIN 166 169
 FT DOMAIN 770 776
 SQ SEQUENCE 793 AA; 91175 MW; 6E063A96620F8C27 CRC64;

Query Match 19.4%; Score 118.5; DB 1; Length 793;
 Best Local Similarity 28.4%; Pred. No. 0.022;
 Matches 38; Conservative 27; Mismatches 36; Indels 33; Gaps 7;

QY 4 PNTTSGSNNVTSGSKNVIA-----GNDNTVIGSDNNVSGSNN-----TVSGND 49
 DB 21 PSSSSPSSNNSTSLKSMISGIENTLVHSKGNKNNNNNNNNNNNNNNNNKOKDIVSLN 80
 QY 50 N-----TVGSNNVTGTHVTNNNNVSGDNNV-SGSFHTVSGHNTVSGS 97
 DB 81 NSSSNTTTTNTTNTTNTSNH--NSNNNSNNNNNNI--NNNNNNNNTVPLVGNHNGCGD 135
 QY 98 --NNTVSGSNHYVS 109
 DB 136 KLNDQPSPSHHVS 149

RESULT 5
 ID N100_YEAST
 AC 002629;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R., Rout M.P., Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 RT NUP100 gene, an open reading frame (ORF) possibly representing a
 RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 RT addition to seven ORFs with weak or no significant similarity to
 RT known proteins";
 RL Yeast 10:S69-S74(1994).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEIN DURING NUCLEOCYTOPLASMIC TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
 CC -----

CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 6 LYSM REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M77639; AAA24776.1; -
CC PIR: A42296; A42296.
CC HSP: P21697; 1PCS.
CC InterPro: IPR002901; Amidase_4.
CC PIRfam: PF01832; Amidase_4; 1.
CC SMART: SM00047; LysM; 6.
CC SMART: SM00257; LysM; 6.
CC Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
CC Cell division; Septation; Repeat; Signal.
CC
CC SIGNAL 1 49
CC CHAIN 50 666 MURAMIDASE-2.
CC REPEAT 257 299 LYSM 1.
CC REPEAT 338 380 LYSM 2.
CC REPEAT 414 456 LYSM 3.
CC REPEAT 489 531 LYSM 4.
CC REPEAT 565 607 LYSM 5.
CC REPEAT 623 665 LYSM 6.
CC SEQUENCE 666 AA; 70670 MW; FFOA7FAFCDB10BA3 CRC64;

Query Match 19.1%; Score 116.5; DB 1; Length 666;
Best Local Similarity 31.0%; Pred. No. 0.026; Mismatches 45; Indels 23; Gaps 5;
Matches 39; Conservative 19;

QY 8 SGSNNTVSGSKNVLAGDNTVIGDNNVSGSNNTVSG-----NDNTVT----- 53
DB 306 SSTVTGNNASGNTSGNT--SGSTGATGAKTYKSGSGSVKINDIGISMNOLIE 363
QY 54 --GSNHYVSGTNIHVDNNNNVSGNDNNV---GSEHTVSGGNTVSGSNNTVSGSNHY 107
DB 364 NNKNNEFYPCQQLVYSKSSASGSGSTSNSTGNTSSNTANTG--STSGSTYTKAGESEV 422

QY 108 VSGSNK 113
DB 423 WSVSNK 428

RESULT 8
PRT3_DICDI STANDARD: PRT: 989 AA.
AC P34637;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Protein-tyrosine phosphatase 3 (EC 3.1.3.48) (Protein-tyrosine-
DE phosphate phosphohydrolase 3).
GN (PTPC1 OR PTP3) AND (PTPC2 OR PTP3).
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96189126; PubMed=8628311;
RA Gampert M., Howard P.K., Hunter T., Firtel R.A.;
RT "Multiple roles of the novel protein tyrosine phosphatase PTP3 during
RT Dictyostellium growth and development."
RL MOL. CELL. BIOL. 16:2431-2444(1996).
CC -1- FUNCTION: SEEMS TO DEPHOSPHORYLATE A PROTEIN OF 130 kDa (P130).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein

CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: IN THE ANTERIOR-LIKE AND PRESTALK CELL TYPES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT MODERATE LEVELS DURING GROWTH
CC AND DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U38197; AAC47041.1; -
CC HSP: Q06124; 2SHP.
CC DictyDB: DD01111; PTPC1.
CC DictyDB: DD02777; PTPC2.
CC InterPro: IPR000387; Tyr_phosphatase.
CC InterPro: IPR000242; Tyr_prot_phphatase.
CC PIRfam: PF00102; Y_phosphatase; 1.
CC PRINTS: PR00700; PTPPHPTASE.
CC SMART: SM00194; PTPC; 1.
CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE: PS0056; TYR_PHOSPHATASE_2; 1.
CC PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 1.
CC Hydrolase.
CC
CC ACT_SITE 649 649
CC DOMAIN 460 716 BY SIMILARITY.
CC DOMAIN 64 71 PROTEIN-TYROSINE PHOSPHATASE.
CC DOMAIN 109 118 POLY-ASN.
CC DOMAIN 137 190 POLY-ASN.
CC DOMAIN 249 257 POLY-SER.
CC DOMAIN 258 265 POLY-THR.
CC DOMAIN 286 289 POLY-ASN.
CC DOMAIN 286 371 POLY-SER.
CC DOMAIN 366 371 POLY-GLN.
CC DOMAIN 787 790 POLY-GLN.
CC DOMAIN 834 839 POLY-GLN.
CC DOMAIN 883 892 POLY-GLN.
CC DOMAIN 906 914 POLY-ASN.
CC DOMAIN 943 963 POLY-ASN.
CC SEQUENCE 989 AA; 109995 MW; 9371105AF80974AF CRC64;

Query Match 19.0%; Score 116; DB 1; Length 989;
Best Local Similarity 28.2%; Pred. No. 0.044;
Matches 29; Conservative 24; Mismatches 40; Indels 10; Gaps 3;

QY 5 NTISG--SNNTVSGSKNVLAGDNTVIGDNNVSGSNNTVSGDNTVTSNHYVSGT 62
DB 96 NSINNKINNTTNNNNNN--NNNNNDKFDTNALSLSTMTIKNNNNNNNNNN----- 147

QY 63 NHTVTDNNNNVSGDNNVSGSEFTVSGGNTVSGSNNTVSGSN 105
DB 148 NNN 190

RESULT 9
YINO_YEAST STANDARD: PRT: 964 AA.
ID YINO_YEAST
AC P40467;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative 108.8 kDa transcriptional regulatory protein in FKH1-STH1
DE intergenic region.
GN YIL130W.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-S288C / AB972;
RA      Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA      Churcher C.M., Connor R., Copsey T., Dear S., Delin K., Fraser A.,
RA      Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA      Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
RA      Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA      Walsh S.V., Whitehead S.;
RL      Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC      -1 SUBCELLULAR LOCATION: Nuclear (Probable)
CC      -1 SIMILARITY: CONTAINS 1 ZN(2)-CY5(6), FUNGAL-TYPE BINUCLEAR
CC      CLUSTER DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: Z38059; CAA86148.1; -
DR      PIR: S48404; S48404.
DR      HSSP: P08657; ICED.
DR      SGD: S0001392; YIL130W.
DR      InterPro: IPR001138; ZN2_Cy6_fungal.
DR      Pfam: PF00172; zn_c1us; 1.
DR      PRINTS: PR00054; FUNGALZNCYS.
DR      SMART: SMO0066; GAI4; 1.
DR      PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR      PROSITE: PS00468; ZN2_Cy6_FUNGAL_2; 1.
DR      KX Hypothetical protein; Transcription regulation; DNA-binding;
DR      KW Nuclear protein; Zinc; Metal-binding.
FT      DNAS_BIND 21 47 ZN(2)-CY5(6), FUNGAL-TYPE.
FT      DOMAIN 811 896 HIS-RICH
SQ      SEQUENCE 964 AA; 108780 MW; AD5ABE59EAB022CC CRC64;

```

Query Match 18.9%; Score 115.5; DB 1; Length 964;
 Best Local Similarity 32.6%; Pred. No. 0.047;
 Matches 29; Conservative 18; Mismatches 37; Indels 5; Gaps 2;

```

OY      16 GSKKNVLAGNDN-TVIGSGDNNSVSGSNNTVYSGNDTVTGSGNHVSGTNIHTVDNNNVS 74
DB      806 SSSKSNSSNNSNPTITNNNNYNNVSGKNN---NNNNITNNSNNNNNNNNNNNNNNNNNN 861
OY      75 GNDNVSGSFTVSGGHNVTGSGSNNTVSG 103
DB      862 NNNNNNNNSGNSNNNNNNNNNNNNNNNDPG 890

```

RESULT 10
 PNT2_DROME STANDARD; PRT; 718 AA.
 AC PS1023; 09VCN2;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ETS-like protein pointed p2 (D-ETS-2).
 GN PNT OR ETS58AB OR ETS2 OR CG17077.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tychcheata; Insecta;
 OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN
 RP SEQUENCE FROM N.A.; ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE=94038653; PubMed=8223245;
 RA Klamant C.;
 RT "The Drosophila gene pointed encodes two ETS-like proteins which are
 RT involved in the development of the midline glial cells.";
 RL Development 117:163-176(1993).

```

RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BERKELEY;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Mortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA      Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA      Baller R.M., Basu A., Bendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA      Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Daveport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA      Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Shen M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shee H.,
RA      Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA      Svizkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).

```

(3)
 RN SEQUENCE OF 551-708 FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL
 RP STAGE.
 RC STRAIN-CANTON-S; TISSUE-Larva;
 RX MEDLINE=92249640; PubMed=1577186;
 RA Chen T., Buntling M., Karim F.D., Thummel C.S.;
 RT "Isolation and characterization of five Drosophila genes that encode
 RT an ets-related DNA binding domain.";
 RL Dev. Biol. 151:176-191(1992).
 CC -1 FUNCTION: REQUIRED FOR GLIAL-NEURONAL CELL INTERACTIONS AT THE
 CC VENTRAL MIDLINE WHICH ARE NECESSARY FOR THE PROPER ELABORATION OF
 CC COMMISSURES IN THE EMBRYONIC CNS.
 CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; P1 (AC PS1022) AND P2 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN A COMPLEX DYNAMIC PATTERN IN
 CC EARLY EMBRYOS, INCLUDING THE MIDLINE AND MIDLINE GLIAL CELLS.
 CC -1 DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH LOWER
 CC LEVELS DURING LARVAL DEVELOPMENT.
 CC -1 SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1 SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X69167; CAA48917.1; -

EMBL: 274215; CAA98741.1; -
DR PIR; S31139; S31139.
DR HSSP; P04170; 6RXN.
DR SGD; S0002326; NRP1.
DR InterPro; IPR000504; RRM.
DR InterPro; IPR001876; Znf-RanBP.
DR Pfam; PF00076; rrm; 1.
DR Pfam; PF00641; zf-RanBP; 2.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; Znf_RBZ; 2.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE; PS01358; ZF_RANBP2_1; 2.
DR PROSITE; PS0199; ZF_RANBP2_2; 2.
DR Nuclear protein; zinc-finger; RNA-binding; Repeat.
FT DOMAIN 226 322
FT ZN_FING 355 384
FT ZN_FING 581 610
FT DOMAIN 490 564
FT CONFLICT 493 493 ASN-RICH.
FT CONFLICT 493 493 I -> N (IN REF. 1).
SQ SEQUENCE 719 AA; 79299 MW; ADA9BC09FD582669 CRC64;

Query Match 17.6%; Score 107.5; DB 1; Length 719;
Best Local Similarity 25.0%; Pred. No. 0.14;
Matches 36; Conservative 21; Mismatches 40; Indels 47; Gaps 6;

QY 5 NTSGSNNTV-----RSGSKNVLAGNDNTVTSGDNNVSGSNNTVSGNDNTVTSGNHV 58
DB 491 NNTNN 546
QY 59 VSGNHTVTDNNNNVSGNDNNVSG-----SFTVTS-----88
DB 547 NSNTNN-----NN 603
QY 89 -GGHNTVSGSNNTVSGSNHVSGS 111
DB 604 CGGPKRSISGD---ASETNHYIDSS 624

RESULT 20
ICEV_PSESX STANDARD; PRT; 1196 AA.
ID ICEV_PSESX
AC 033479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAV.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INAV;
RX MEDLINE=97462815; PubMed=9323042;
RA Schmid D., Primrose D., Capitani G., Battistutta R., Neeser J.-R.,
Jann A.;
RT "Molecular organisation of the ice nucleation protein Inav from
Pseudomonas syringae.";
RT FEBS Lett. 414:590-594(1997).
RL -1- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -1- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
FOR ICE NUCLEATION ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001086; CAA04521.1; -
DR HSSP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 61.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 42.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 172 1147
FT CONFLICT 172 1147 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1196 AA; 117991 MW; C9E9974CB1731E68 CRC64;

Query Match 17.6%; Score 107.5; DB 1; Length 1196;
Best Local Similarity 27.6%; Pred. No. 0.25;
Matches 37; Conservative 22; Mismatches 48; Indels 27; Gaps 6;

QY 9 GSNNTVSGSKNVLAGNDNTVTSG-DNNVSGSNNTVSGNDNTVTS-----GSN 56
DB 229 GSTGTARBSG-NLFTAGYSTGTAGSDSLIAGYSTGTSGSDSLIAGYSTGTARBSG 297
QY 57 HVT-----SGNHTVTDNNNNVSGNDNNVSG-----SFTVTS-----102
DB 298 LTAGYSTGTAGSDSLIAGYSTGTSGSDSLIAGYSTGTARBSGSLIAGYSTGTAG 357
QY 103 GSNHVSGSNKVT 116
DB 358 ADSLIAGYSTGT 371

RESULT 21
SWI1_YEAST STANDARD; PRT; 1314 AA.
ID SWI1_YEAST
AC P09547;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription regulatory protein Swi1 (SWI/SNF complex component SWI1)
DE (Transcription regulatory protein ADR6) (Regulatory protein GAN3).
GN ADR6 OR SWI1 OR GAN3 OR YPL016W OR IPA1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89057455; PubMed=3143101;
RA O'Hara P.J., Horowitz H., Eichinger H., Young E.T.;
RT "The yeast ADR6 gene encodes homopolymeric amino acid sequences and a
potential metal-binding domain.";
RT Nucleic Acids Res. 16:10153-10170(1988).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=92154671; PubMed=1339306;
RA Peterson C.L., Herskowitz I.;
RT "Characterization of the yeast Swi1, Swi2, and Swi3 genes, which
encode a global activator of transcription.";
RL Cell 68:573-583(1992).
CC -1- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC -1- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
COMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 ARID DOMAIN.

RP SEQUENCE FROM N.A.
RC STRAIN-5288C;
RA MEDLINE=97435481; PubMed=9290212;
R Rieger M., Bruckner M., Schaefer M.,
RT "Sequence analysis of 203 kilobases from *Saccharomyces cerevisiae*
RT chromosome VII.",
LT Yeast 13:1077-1090(1997).

RA Chen T., Bunting M., Katim F.D., Thummel C.S.:
RT "Isolation and Characterization of five Drosophila genes that encode
RT an esr-related DNA binding domain.",
RL Dev. Biol. 151,176-191(1992).
RN [3]
RN
RN SEQUENCE OF 445-603 FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.

CC Hypocreales: Nectriaceae; Gibberella.
 OX NCBI_TaxID=5127;
 RN
 RP SEQUENCE FROM N.A.
 RA Capriati C., Richter A., Bergmann C., Lo Cicero S., Salvi G.,
 RT Cervone F., de Lorenzo G.;
 RT "Cloning and characterization of a gene encoding the
 RT endopolygalacturonase of Fusarium moniliforme";
 RL Mycol. Res. 97:497-505(1993).
 CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
 CC galactosiduronic linkages in pectate and other galacturonans.
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
 CC (POLYGALACTURONASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: I02239; AAA74586.1; -
 DR InterPro: IPR000743; Polygalacturonase.
 DR Pfam: PF00295; Glyco_hydro_28; 1.
 DR PROSITE: PS00502; POLYGALACTURONASE; 1.
 KW Hydrolyase; Glycosidase; Cell wall; Signal; Zymogen; Glycoprotein.
 FT SIGNAL 1 24
 FT CHAIN 1 25
 FT ACT_SITE 234 234
 FT CARBOHYD 65 65 BY SIMILARITY.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 373 AA; 38915 MW; D77BCACAC7FFBD8F CRC64;
 SO
 Query Match 17.3%; Score 105.5; DB 1; Length 373;
 Best local similarity 24.3%; Pred. No. 0.099;
 Matches 44; Conservative 22; Mismatches 46; Indels 69; Gaps 8;
 QY 1 DEQPTTIGSNNTVRSGSKNVLGNDNTVISG---DNNS----- 36
 DB 84 DENPIVIGSNNTITGASGHVIDGNGQAYWDGKGSNSNSNOKPDHFIYVQKTGNSKITN 143
 QY 37 -----VSGSNTTVSG-----NDNTVTS-----NHVSGTNHIYT 67
 DB 144 INIQMWPVHCEDITGSSQLTITISGLILDNKAGDKPKNAKSGSLPAHNTDGFIDISSDHYTL 203
 QY 68 DNNNNVSGND-----NNVSGSFHTVSGH---NTVSG-SNNTVSG---SNHVSGS 111
 DB 204 DNNHYNDQDCVAVTSGNIVVSNMYCSGGHLSIGSYGKSDNVVDGVFLSSQVNSQ 263
 QY 112 N 112
 DB 264 N 264

Search completed: July 29, 2002, 06:50:58
 Job time: 92 sec